THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

VOLUME XII

1941

311
ANN.

THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG J. NEYMAN

WITH THE COÖPERATION OF

H. C. CARVER
H. CRAMÉR
W. E DEMING
G. DARMOIS
R. A. FISHER
T. C. FRY
H. HOTELLING
R. DE MISES
E. S. PEARSON
H. L. RIETZ
W. A. SHEWHART

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4 00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I–IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each Single numbers $1.25.

Subscriptions, renewals, orders for back numbers and other business communications should be sent to A. T. Craig, University of Iowa, Iowa City, Iowa.

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

CONTENTS OF VOLUME XII

Articles

Aroian, Leo A. Continued Fractions for the Incomplete Beta Function. . 218

Aroian, Leo A. A Study of R. A. Fisher's z-Distribution and the Related F-Distribution 429

Beale, Frank S. On a Certain Class of Orthogonal Polynomials 97

Bellinson, H. R., von Neumann, J., Kent, R. H., and Hart, B. I. The Mean Square Successive Difference. 153

Brookner, R. J. and Wald, A On the Distribution of Wilks' Statistic for Testing the Independence of Several Groups of Variates . 137

Chung, Kai Lai. On the Probability of the Occurrence of at Least m Events Among n Arbitrary Events 328

Curtiss, J. H. On the Distribution of the Quotient of Two Chance Variables 409

Dodd, Edward L. Some Generalizations of the Logarithmic Mean and of Similar Means of Two Variates Which Become Indeterminate When the Two Variates are Equal 422

Dodd, Edward L. The Cyclic Effects of Linear Graduations Persisting in the Differences of the Graduated Values 127

Doob, J. L. and von Mises, R. Discussion of Papers on Probability Theory 215

Doob, J. L. Probability as Measure 206

Dressel, Paul L A Symmetric Method for Obtaining Unbiased Estimates and Expected Values 84

Dwyer, Paul S. The Doolittle Technique 449

Dwyer, P. S. The Skewness of the Residuals in Linear Regression Theory 104

Feller, Willy. On the Integral Equation of Renewal Theory. 243

Gumbel, E. J. The Return Period of Flood Flows. 163

Hart, B. I., von Neumann, J., Kent, R. H., and Bellinson, H. R. The Mean Square Successive Difference 153

Hotelling, Harold. Experimental Determination of the Maximum of a Function. 20

Hsu, Chung Tsi. Samples from Two Bivariate Normal Populations. . 279

Kent, R. H., von Neumann, J., Bellinson, H. R., and Hart, B. I The Mean Square Successive Difference. 153

McPherson, J. C. On Mechanical Tabulation of Polynomials 317

Mood, A. M. On the Joint Distribution of the Medians in Samples from a Multivariate Population. 268

NEYMAN, J. On a Statistical Problem Arising in Routine Analyses and in Sampling Inspections of Mass Production 46
PAULSON, EDWARD. On Certain Likelihood-Ratio Tests Associated with the Exponential Distribution 301
SATTERTHWAITE, FRANKLIN E. A Concise Analysis of Certain Algebraic Forms 77
TUCKERMAN, L. B. On the Mathematically Significant Figures in the Solution of Simultaneous Linear Equations 307
VON MISES, R. AND DOOB, J. L. Discussion of Papers on Probability Theory 215
VON MISES, R. On the Foundations of Probability and Statistics . . . 191
VON NEUMANN, JOHN. Distribution of the Ratio of the Mean Successive Difference to the Variance 367
VON NEUMANN, J., KENT, R. H., BELLINSON, H R., AND HART, B. I. The Mean Square Successive Difference 153
WALD, ABRAHAM Asymptotically Most Powerful Tests of Statistical Hypotheses 1
WALD, ABRAHAM Some Examples of Asymptotically Most Powerful Tests 396
WALD, A. AND BROOKNER, R. J. On the Distribution of Wilks' Statistic for Testing the Independence of Several Groups of Variates 137
WILKS, S. S. On the Determination of Sample Sizes for Setting Tolerance Limits 91
YOUNG, L C On Randomness in Ordered Sequences 293

NOTES

BAKER, G. A. Tests of Homogeneity for Normal Populations 233
CRAIG, CECIL C. A Note on Sheppard's Correction 339
CRAIG, CECIL C. Note on the Distribution of Non-central t with an Application 224
DALY, JOSEPH F. A Problem in Estimation 459
GORDON, ROBERT D. The Estimation of a Quotient when the Denominator is Normally Distributed 115
GORDON, ROBERT D. Values of Mills' Ratio of Area to Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument 364
GREVILLE, T. N. E. The Frequency Distribution of a General Matching Problem 350
HOEL, PAUL G. On Methods of Solving Normal Equations 354
KAVANAGH, ARTHUR J. Note on the Adjustment of Observations . . . 111
KENDALL, M G Corrections to a Paper on the Uniqueness Problem of Moments 464
KOLMOGOROFF, A. Confidence Limits for an Unknown Distribution Function 461

MOSTELLER, FREDERICK Note on an Application of Runs to Quality Control Charts 228
SAMUELSON, PAUL A. Conditions that the Roots of a Polynomial be Less Than Unity in Absolute Value 360
STEWART, W MAC A Note on the Power of the Sign Test 236
WALD, ABRAHAM On the Analysis of Variance in Case of Multiple Classifications with Unequal Class Frequencies 346
WALD, A. AND WOLFOWITZ, J. Note on Confidence Limits for Continuous Distribution Functions 118
WILLIAMS, J. D. Moments of the Ratio of the Mean Square Successive Difference to the Mean Square Difference in Samples from a Normal Universe 239
WOLFOWITZ, J. AND WALD, A. Note on Confidence Limits for Continuous Distribution Functions 118

MISCELLANEOUS

Abstracts of Papers 123, 470
Announcement Concerning Computation of Mathematical Tables. . . . 465
Constitution and By-Laws of the Institute 474
Directory of the Institute 478
Report of the Chicago Meeting of the Institute (Dec. 1940) 120
Report of the Chicago Meeting of the Institute (Sept. 1941) 468

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

PAGE

Asymptotically Most Powerful Tests of Statistical Hypotheses. ABRAHAM WALD 1

Experimental Determination of the Maximum of a Function. HAROLD HOTELLING 20

On a Statistical Problem Arising in Routine Analyses and in Sampling Inspections of Mass Production. J. NEYMAN 40

A Concise Analysis of Certain Algebraic Forms. FRANKLIN E. SATTERTHWAITE 77

A Symmetric Method of Obtaining Unbiased Estimates and Expected Values. PAUL L. DRESSEL 84

Determination of Sample Sizes for Setting Tolerance Limits. S. S. WILKS 91

On a Certain Class of Orthogonal Polynomials. FRANK S. BEALE 97

The Skewness of the Residuals in Linear Regression Theory. P. S. DWYER 104

Notes:

Note on the Adjustment of Observations. ARTHUR J. KAVANAGH 111

The Estimation of a Quotient When the Denominator is Normally Distributed. ROBERT D. GORDON 115

Note on Confidence Limits for Continuous Distribution Functions. A. WALD AND J. WOLFOWITZ 118

Report of the Chicago Meeting of the Institute 120

Abstracts of Papers 123

Vol. XII, No. 1 — March, 1941

THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG J. NEYMAN

WITH THE COÖPERATION OF

H. C. CARVER
H. CRAMÉR
W. E. DEMING
G. DARMOIS
R. A. FISHER
T. C. FRY
H. HOTELLING
R. DE MISES
E. S. PEARSON
H. L. RIETZ
W. A. SHEWHART

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I-IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

ASYMPTOTICALLY MOST POWERFUL TESTS OF STATISTICAL HYPOTHESES[1]

By Abraham Wald[2]

Columbia University, New York City

1. Introduction. Let $f(x, \theta)$ be the probability density function of a variate x involving an unknown parameter θ. For testing the hypothesis $\theta = \theta_0$ by means of n independent observations $x_1, \cdots, x_n$ on x we have to choose a region of rejection W_n in the n-dimensional sample space. Denote by $P(W_n \mid \theta)$ the probability that the sample point $E = (x_1, \cdots, x_n)$ will fall in W_n under the assumption that θ is the true value of the parameter. For any region U_n of the n-dimensional sample space denote by $g(U_n)$ the greatest lower bound of $P(U_n \mid \theta)$. For any pair of regions U_n and T_n denote by $L(U_n, T_n)$ the least upper bound of

$$P(U_n \mid \theta) - P(T_n \mid \theta).$$

In all that follows we shall denote a region of the n-dimensional sample space by a capital letter with the subscript n.

Definition 1. A sequence $\{W_n\}$, $(n = 1, 2, \cdots, \text{ad inf.})$, of regions is said to be an asymptotically most powerful test of the hypothesis $\theta = \theta_0$ on the level of significance α if $P(W_n \mid \theta_0) = \alpha$ and if for any sequence $\{Z_n\}$ of regions for which $P(Z_n \mid \theta_0) = \alpha$, the inequality

$$\limsup_{n \to \infty} L(Z_n, W_n) \leq 0$$

holds.

Definition 2. A sequence $\{W_n\}$, $(n = 1, 2, \cdots, \text{ad inf.})$, of regions is said to be an asymptotically most powerful unbiased test of the hypothesis $\theta = \theta_0$ on the level of significance α if $P(W_n \mid \theta_0) = \lim_{n=\infty} g(W_n) = \alpha$, and if for any sequence $\{Z_n\}$ of regions for which $P(Z_n \mid \theta_0) = \lim_{n=\infty} g(Z_n) = \alpha$, the inequality

$$\limsup_{n \to \infty} L(Z_n, W_n) \leq 0$$

holds.

Let $\hat{\theta}_n(x_1, \cdots, x_n)$ be the maximum likelihood estimate of θ in the n-dimensional sample space. That is to say, $\hat{\theta}_n(x_1, \cdots, x_n)$ denotes the value of θ

[1] Presented to the American Mathematical Society at New York, February 24, 1940.

[2] Research under a grant-in-aid from the Carnegie Corporation of New York.

for which the product $\prod_{\nu=1}^{n} f(x_\nu, \theta)$ becomes a maximum. Let W_n' be the region defined by the inequality $\sqrt{n}(\hat{\theta}_n - \theta_0) \geq c_n'$, W_n'' defined by the inequality $\sqrt{n}(\hat{\theta}_n - \theta_0) \leq c_n''$, and let W_n consists of all points for which at least one of the inequalities

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \geq a_n, \qquad \sqrt{n}(\hat{\theta}_n - \theta_0) \leq -a_n$$

is satisfied. The constants a_n, c_n', c_n'' are chosen such that

$$P(W_n' \mid \theta_0) = P(W_n'' \mid \theta_0) = P(W_n \mid \theta_0) = \alpha.$$

It will be shown in this paper that under certain restrictions on the probability density $f(x, \theta)$ the sequence $\{W_n'\}$ is an asymptotically most powerful test of the hypothesis $\theta = \theta_0$ if θ takes only values $\theta \geq \theta_0$. Similarly $\{W_n''\}$ is an asymptotically most powerful test if θ takes only values $\theta \leq \theta_0$. Finally $\{W_n\}$ is an asymptotically most powerful unbiased test if θ can take any real value.

2. Assumptions on the density function $f(x, \theta)$.

ASSUMPTION 1. *For any positive k*

$$\lim_{n=\infty} P(-k < \hat{\theta}_n - \theta < k \mid \theta) = 1$$

uniformly in θ, where $P(-k < \hat{\theta}_n - \theta < k \mid \theta)$ denotes the probability that $-k \leq \hat{\theta}_n - \theta \leq k$ under the assumption that θ is the true value of the parameter.

Assumption 1 implies somewhat more than consistency of the maximum likelihood estimate $\hat{\theta}_n$. In fact, consistency means only that for any positive k

$$\lim_{n=\infty} P(-k \leq \hat{\theta}_n - \theta \leq k \mid \theta) = 1,$$

without asking that the convergence should be uniform in θ. If $\hat{\theta}_n$ satisfies Assumption 1 we shall say that $\hat{\theta}_n$ is a uniformly consistent estimate of θ. A rigorous proof of the consistency of $\hat{\theta}_n$ (under certain restrictions on $f(x, \theta)$) was given by J. L. Doob.[3] In an appendix to this paper it will be shown that under certain conditions $\hat{\theta}_n$ is uniformly consistent.

Denote by $E_\theta[\psi(x)]$ the expected value of $\psi(x)$ under the assumption that θ is the true value of the parameter. That is to say,

$$E_\theta[\psi(x)] = \int_{-\infty}^{\infty} \psi(x) f(x, \theta)\, dx.$$

For any x, for any positive δ, and for any θ_1, denote by $\varphi_1(x, \theta_1, \delta)$ the greatest lower bound, and by $\varphi_2(x, \theta_1, \delta)$ the least upper bound of $\dfrac{\partial^2 \log f(x, \theta)}{\partial \theta^2}$ in the interval $\theta_1 - \delta \leq \theta \leq \theta_1 + \delta$.

ASSUMPTION 2. *There exists a positive value k_0 such that the expectations $E_\theta \varphi_1(x, \theta_1, \delta)$ and $E_\theta \varphi_2(x, \theta_1, \delta)$ exist and are continuous functions of θ, θ_1 and δ*

[3] J. L. Doob, "Probability and statistics," *Trans. Am. Math Soc.*, Vol. 36 (1937).

in the domain D defined by the inequalities: $0 \leq \delta \leq \frac{1}{2}k_0$, $\theta_0 - \frac{1}{2}k_0 \leq \theta_1 \leq \theta_0 + \frac{1}{2}k_0$, $\theta_0 - k_0 \leq \theta \leq \theta_0 + k_0$. *Furthermore the expectations* $E_\theta[\varphi_1(x, \theta_1, \delta)]^2$ *and* $E_\theta[\varphi_2(x, \theta_1, \delta)]^2$ *exist in D and have a finite upper bound in D.*

ASSUMPTION 3. *There exists a positive value k_0 such that*

$$\int_{-\infty}^{\infty} \frac{\partial f(x, \theta)}{\partial \theta}\, dx = \int_{-\infty}^{\infty} \frac{\partial^2 f(x, \theta)}{\partial \theta^2}\, dx = 0 \quad \text{for} \quad \theta_0 - k_0 \leq \theta \leq \theta_0 + k_0.$$

Assumption 3 means simply that we may differentiate with respect to θ under the integral sign. In fact

$$\int_{-\infty}^{\infty} f(x, \theta)\, dx = 1$$

identically in θ. Hence

$$\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(x, \theta)\, dx = \frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{\infty} f(x, \theta)\, dx = 0.$$

Differentiating under the integral sign, we obtain the relations in Assumption 3.

ASSUMPTION 4 *There exists a positive η and a positive k_0 such that*

$$E_\theta \left| \frac{\partial \log f(x, \theta)}{\partial \theta} \right|^{2+\eta}$$

exists and has a finite upper bound in the interval $\theta_0 - k_0 \leq \theta \leq \theta_0 + k_0$.

3. Some propositions. Denote $\sqrt{n}\,(\theta_n - \theta)$ by $z_n(\theta)$ and denote the probability $P[z_n(\theta) < t \mid \theta]$ by $\Phi_n(t, \theta)$.

PROPOSITION I. *Within the θ-interval* $[\theta_0 - \frac{1}{2}k_0, \theta_0 + \frac{1}{2}k_0]$ $\Phi_n(t, \theta)$ *converges with* $n \to \infty$ *uniformly in t and θ towards the cumulative normal distribution with zero mean and variance*

$$-1 \Big/ E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2}$$

PROOF: In all that follows we assume that θ takes only values in the interval $[\theta_0 - k_0, \theta_0 + k_0]$, except when the contrary is explicitly stated. Furthermore we introduce the variable θ_1 and assume that θ_1 takes only values in the interval $[\theta_0 - \frac{1}{2}k_0, \theta_0 + \frac{1}{2}k_0]$.

Because of Assumption 3 we have

$$E_\theta \frac{\partial \log f(x, \theta)}{\partial \theta} = \int_{-\infty}^{\infty} \frac{\partial f(x, \theta)}{\partial \theta}\, dx = 0 \tag{1}$$

Since

$$\frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} = \frac{1}{f(x, \theta)} \cdot \frac{\partial^2 f(x, \theta)}{\partial \theta^2} - \frac{1}{[f(x, \theta)]^2} \left[\frac{\partial f(x, \theta)}{\partial \theta} \right]^2$$

we get from Assumption 3

$$E_\theta \left[\frac{\partial \log f(x, \theta)}{\partial \theta} \right]^2 = -E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2}. \tag{2}$$

Hence

$$d(\theta) = -E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} > 0. \tag{3}$$

Consider the Taylor expansion

$$\sum_{\alpha=1}^{n} \frac{\partial \log f(x_\alpha, \theta)}{\partial \theta} = \sum_{\alpha=1}^{n} \frac{\partial \log f(x_\alpha, \theta_1)}{\partial \theta} + (\theta - \theta_1) \sum_{\alpha=1}^{n} \frac{\partial^2 \log f(x_\alpha, \theta')}{\partial \theta^2} \tag{4}$$

where θ' lies in the interval $[\theta_1, \theta]$. Denote $\frac{1}{\sqrt{n}} \sum_\alpha \frac{\partial \log f(x_\alpha, \theta_1)}{\partial \theta}$ by $y_n(\theta_1)$. For $\theta = \hat{\theta}_n$ the left hand side of (4) is equal to zero. Hence we have

$$y_n(\theta_1) + [\sqrt{n}(\hat{\theta}_n - \theta_1)] \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \theta')}{\partial \theta^2} = 0, \tag{5}$$

or

$$y_n(\theta_1) + z_n(\theta_1) \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \theta')}{\partial \theta^2} = 0. \tag{6}$$

Let $Q_n(\theta_1)$ be the region defined by the inequality

$$\left| \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \theta')}{\partial \theta^2} + d(\theta_1) \right| < \nu \tag{7}$$

where ν denotes a positive number less than the greatest lower bound of $d(\theta_1)$. We shall prove that

$$\lim_{n=\infty} P[Q_n(\theta_1) \mid \theta_1] = 1 \tag{8}$$

uniformly in θ_1. Let τ_0 be a positive number such that

$$\left| E_{\theta_1} \varphi_i(x, \theta_1, \tau_0) - E_{\theta_1} \frac{\partial^2 \log f(x, \theta_1)}{\partial \theta^2} \right| < \frac{\nu}{2}, \qquad (i = 1, 2) \tag{9}$$

for all values of θ_1. Because of Assumption 2 such a τ_0 certainly exists. Denote by $R_n(\theta_1)$ the region defined by the inequality

$$| \hat{\theta}_n - \theta_1 | \leq \tau_0 . \tag{10}$$

On account of Assumption 1

$$\lim_{n=\infty} P[R_n(\theta_1) \mid \theta_1] = 1 \tag{11}$$

uniformly in θ_1. Since θ' lies in the interval $[\theta_1, \hat{\theta}_n]$, we have

$$| \theta' - \theta_1 | \leq \tau_0 \tag{12}$$

for all points in $R_n(\theta_1)$. Hence at any point in $R_n(\theta_1)$ the inequality

$$\sum_{\alpha=1}^{n} \varphi_1(x_\alpha, \theta_1, \tau_0) \leq \sum_{\alpha=1}^{n} \frac{\partial^2 \log f(x_\alpha, \theta')}{\partial \theta^2} \leq \sum_{\alpha=1}^{n} \varphi_2(x_\alpha, \theta_1, \tau_0) \tag{13}$$

holds.

Let $S_n(\theta_1)$ be defined by the inequality

$$\left|\frac{1}{n}\sum_{\alpha}\varphi_1(x_\alpha, \theta_1, \tau_0) - E_{\theta_1}\varphi_1(x, \theta_1, \tau_0)\right| < \frac{\nu}{2}. \tag{14}$$

and $T_n(\theta_1)$ by the inequality

$$\left|\frac{1}{n}\sum_{\alpha}\varphi_2(x_\alpha, \theta_1, \tau_0) - E_{\theta_1}\varphi_2(x, \theta_1, \tau_0)\right| < \frac{\nu}{2}. \tag{15}$$

On account of Assumption 2 we have

$$\lim_{n=\infty} P[S_n(\theta_1) \mid \theta_1] = \lim_{n=\infty} P[T_n(\theta_1) \mid \theta_1] = 1 \tag{16}$$

uniformly in θ_1.
Denote by $U_n(\theta_1)$ the common part of the regions $R_n(\theta_1)$, $S_n(\theta_1)$ and $T_n(\theta_1)$. In $U_n(\theta_1)$ we have on account of (9), (14) and (15)

$$\left|\frac{1}{n}\sum_{\alpha}\varphi_i(x_\alpha, \theta_1, \tau_0) - E_{\theta_1}\frac{\partial^2 \log f(x, \theta_1)}{\partial\theta^2}\right| < \nu \qquad (i = 1, 2). \tag{17}$$

From this we obtain (7) because of (13). That is to say, the inequality (7) is valid everywhere in $U_n(\theta_1)$. Since

$$\lim_{n=\infty} P[U_n(\theta_1) \mid \theta_1] = 1$$

uniformly in θ_1, our statement about $Q_n(\theta_1)$ is proved. From (6) and (7) we get that everywhere in $Q_n(\theta_1)$ the inequalities hold:

$$\frac{y_n(\theta_1)}{d(\theta_1) + \nu} \leq z_n(\theta_1) \leq \frac{y_n(\theta_1)}{d(\theta_1) - \nu} \quad \text{if} \quad y_n(\theta_1) \geq 0; \tag{18}$$

$$\frac{y_n(\theta_1)}{d(\theta_1) + \nu} \geq z_n(\theta_1) \geq \frac{y_n(\theta_1)}{d(\theta_1) - \nu} \quad \text{if} \quad y_n(\theta_1) \leq 0. \tag{19}$$

Let $z_n^*(\theta_1)$ be defined as follows: $z_n^*(\theta_1) = z_n(\theta_1)$ at any point in $Q_n(\theta_1)$, and $z_n^*(\theta_1) = y_n(\theta_1)/d(\theta_1)$ at any point outside $Q_n(\theta_1)$.
On account of (8) we obviously have

$$\lim_{n=\infty} P[z_n^*(\theta_1) < t \mid \theta_1] - P[z_n(\theta_1) < t \mid \theta_1] = 0 \tag{20}$$

uniformly in t and θ_1.
From equation (1) it follows that $E_{\theta_1} y_n(\theta_1) = 0$. From Assumption 4 it follows on account of the general limit theorems

$$\lim_{n=\infty} P[y_n(\theta_1) < t \mid \theta_1] - \frac{1}{\sqrt{2\pi d(\theta_1)}}\int_{-\infty}^{t} e^{-\frac{1}{2}t^2/d(\theta_1)}\,dt = 0 \tag{21}$$

uniformly in t and θ_1. Hence

$$\lim_{n=\infty} P\left[\frac{y_n(\theta_1)}{d(\theta_1)} < t \mid \theta_1\right] - \sqrt{\frac{d(\theta_1)}{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}t^2 d(\theta_1)}\,dt = 0$$

uniformly in t and θ_1. Since ν can be chosen arbitrarily small, we get easily from (18), (19), (20) and (21)

$$\lim_{n=\infty} \left| P\left[\frac{y_n(\theta_1)}{d(\theta_1)} < t \mid \theta_1\right] - P[z_n(\theta_1) < t \mid \theta_1] \right| = 0 \tag{22}$$

uniformly in t and θ_1. Proposition 1 follows from (21) and (22).

PROPOSITION 2. *Let $\{W_n\}$ be a sequence of regions of size α, i.e. $P(W_n \mid \theta_0) = \alpha$, and let $V_n(z)$ be the region defined by the inequality*

$$(\hat{\theta}_n - \theta_0)\sqrt{n} < z.$$

Let $U_n(z)$ be the intersection of $V_n(z)$ and W_n, and denote $P[U_n(z) \mid \theta_0]$ by $F_n(z)$. Denote furthermore $P[W_n \mid \theta_0 + \mu/\sqrt{n}]$ by $G(\mu, n)$. If $F_n(z)$ converges to $F(z)$ and if $\lim_{n=\infty} \mu_n = \mu$, then

$$\lim_{n=\infty} G(\mu_n, n) = \int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - 2\mu z)/c}\, dF(z) \tag{23}$$

where

$$c = -1 \Big/ E_{\theta_0} \frac{\partial^2 \log f(x, \theta_0)}{\partial \theta^2}.$$

PROOF: First we show

$$\int_{-\infty}^{\infty} dF(z) = \alpha. \tag{24}$$

Denote $P[V_n(z) \mid \theta_0]$ by $\Phi_n(z)$. On account of Proposition 1 $\Phi_n(z)$ converges uniformly to the cumulative normal distribution $\psi(z)$ with zero mean and variance c. It is obvious that

$$F_n(z_2) - F_n(z_1) \leq \Phi_n(z_2) - \Phi_n(z_1) \text{ for } z_2 > z_1. \tag{25}$$

Hence

$$F(z_2) - F(z_1) \leq \psi(z_2) - \psi(z_1) \text{ for } z_2 > z_1. \tag{26}$$

From (25) we get

$$\left[\lim_{z=\infty} F_n(z)\right] - F_n(z) = \alpha - F_n(z) \leq 1 - \Phi_n(z). \tag{27}$$

Hence

$$\alpha - F(z) \leq 1 - \psi(z). \tag{28}$$

Since $F_n(z) \leq \alpha$ and therefore also $F(z) \leq \alpha$, we get from (28)

$$0 \leq \alpha - F(z) \leq 1 - \psi(z).$$

Hence

$$\lim_{z=\infty} F(z) = \alpha. \tag{29}$$

Since $F_n(z) \leq \Phi_n(z)$, we have $F(z) \leq \psi(z)$, and therefore

$$\lim_{z=-\infty} F(z) = 0. \tag{30}$$

The equation (24) follows from (29) and (30).
It follows easily from (26) that the integral on the right hand side of the equation (23) exists and is finite.
Let us denote $\theta_0 + \mu_n/\sqrt{n}$ by θ_n. Consider the Taylor expansions

$$\sum_\alpha \log f(x_\alpha, \theta_0) = \sum_\alpha \log f(x_\alpha, \hat{\theta}_n) + (\theta_0 - \hat{\theta}_n) \sum_\alpha \frac{\partial}{\partial\theta} \log f(x_\alpha, \hat{\theta}_n) + \tfrac{1}{2}(\theta_0 - \hat{\theta}_n)^2 \sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta_n') \tag{31}$$

and

$$\sum_\alpha \log f(x_\alpha, \theta_n) = \sum_\alpha \log f(x_\alpha, \hat{\theta}_n) + (\theta_n - \hat{\theta}_n) \sum_\alpha \frac{\partial}{\partial\theta} \log f(x_\alpha, \hat{\theta}_n) + \tfrac{1}{2}(\theta_n - \hat{\theta}_n)^2 \sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta_n'') \tag{32}$$

where θ_n' lies in the interval $[\theta_0, \hat{\theta}_n]$ and θ_n'' lies in the interval $[\theta_n, \hat{\theta}_n]$. Since $\hat{\theta}_n$ is the maximum likelihood estimate, we get from (31) and (32)

$$\sum_\alpha \log f(x_\alpha, \theta_0) = \sum_\alpha \log f(x_\alpha, \hat{\theta}_n) + \tfrac{1}{2}(\theta_0 - \hat{\theta}_n)^2 \sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta_n'), \tag{33}$$

$$\sum_\alpha \log f(x_\alpha, \theta_n) = \sum_\alpha \log f(x_\alpha, \hat{\theta}_n) + \tfrac{1}{2}(\theta_n - \hat{\theta}_n)^2 \sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta_n''). \tag{34}$$

Denote by β a real variable which can take any value between -2μ and $+2\mu$. Denote by R_n the region defined by the inequality

$$|\hat{\theta}_n - \theta_0| < n^{-\frac{1}{4}}. \tag{35}$$

From Proposition 1 it follows easily that

$$\lim_{n=\infty} P(R_n \mid \theta_0 + \beta/\sqrt{n}) = 1 \tag{36}$$

uniformly in β. Denote $2n^{-\frac{1}{4}}$ by τ_n. Then for almost all n the following inequalities hold at any point in R_n:

$$\sum_\alpha \varphi_1(x_\alpha, \theta_0, \tau_n) \leq \sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta_n') \leq \sum_\alpha \varphi_2(x_\alpha, \theta_0, \tau_n), \tag{37}$$

$$\sum_\alpha \varphi_1(x_\alpha, \theta_0, \tau_n) \leq \sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta_n'') \leq \sum_\alpha \varphi_2(x_\alpha, \theta_0, \tau_n). \tag{38}$$

Denote by S_n the region in which (35), (37) and (38) simultaneously hold. It is obvious that

$$\lim_{n=\infty} P(S_n \mid \theta_0 + \beta/\sqrt{n}) = 1$$

uniformly in β. Denote $\theta_0 + \beta/\sqrt{n}$ by $\theta_n(\beta)$. From Assumption 2 it follows easily that

$$\lim_{n=\infty} E_{\theta_n(\beta)} \left\{ \frac{\sum_\alpha \varphi_i(x_\alpha, \theta_0, \tau_n)}{n} \right\} = E_{\theta_0} \frac{\partial^2}{\partial\theta^2} \log f(x, \theta_0) = \frac{-1}{c} \quad (i = 1, 2) \tag{39}$$

uniformly in β. Furthermore the variance of $\sum_\alpha \frac{\varphi_i(x_\alpha, \theta_0, \tau_n)}{n}$, if $\theta_n(\beta)$ is the true value of the parameter θ, converges to zero with $n \to \infty$ uniformly in β. Hence a sequence $\{\lambda_n\}$, $(n = 1, 2, \cdots, \text{ad inf.})$, of positive numbers can be given such that

$$\lim_{n=\infty} \lambda_n = 0 \tag{40}$$

and

$$\lim P[T_n \mid \theta_n(\beta)] = 1 \tag{41}$$

uniformly in β, where the region T_n is defined by the inequality

$$\left| \sum_\alpha \frac{\varphi_i(x_\alpha, \theta_0, \tau_n)}{n} + \frac{1}{c} \right| < \lambda_n n^{-\frac{1}{2}} \qquad (i = 1, 2). \tag{42}$$

From (37) and (38) it follows that in the intersection T_n' of T_n and S_n

$$\left| \frac{1}{n} \sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta_n') + \frac{1}{c} \right| < \lambda_n n^{-\frac{1}{2}} \tag{43}$$

and

$$\left| \frac{1}{n} \sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta_n'') + \frac{1}{c} \right| < \lambda_n n^{-\frac{1}{2}}. \tag{44}$$

We get from (33), (34), (35), (43) and (44) that at any point in T_n'

$$\sum_\alpha \log f(x_\alpha, \theta_n) - \sum_\alpha \log f(x_\alpha, \theta_0) = \frac{n}{2c} [(\theta_0 - \hat{\theta}_n)^2 - (\theta_n - \hat{\theta}_n)^2] + \lambda_n', \tag{45}$$

where $|\lambda_n'| \leq \rho\lambda_n$, and ρ denotes a constant not depending on n.
On account of (36) and (41) we have

$$\lim_{n=\infty} P[T_n' \mid \theta_n(\beta)] = 1 \tag{46}$$

uniformly in β.
Denote by $T_n''(z)$ the intersection of $U_n(z)$ (defined in Proposition 2) and T_n'. Denote furthermore $P[T_n''(z) \mid \theta_0]$ by $F_n^*(z)$.
Since

$$n[(\theta_0 - \hat{\theta}_n)^2 - (\theta_n - \hat{\theta}_n)^2] = n[(\theta_0 - \hat{\theta}_n)^2 - (\theta_0 - \hat{\theta}_n + \mu_n/\sqrt{n})^2]$$
$$= -\mu_n^2 + 2\sqrt{n}\,\mu_n(\hat{\theta}_n - \theta_0),$$

we get from (45) and (46)

$$\lim_{n\to\infty}\left\{P[T_n''(z)\,|\,\theta_n] - \int_{-\infty}^{z} e^{-\frac{1}{2}(\mu_n^2-2\mu_n t)/c}\,dF_n^*(t)\right\} = 0 \tag{47}$$

uniformly in z. It is obvious that

$$\lim_{n\to\infty}\{P[T_n''(z)\,|\,\theta_n] - P[U_n(z)\,|\,\theta_n]\} = 0 \tag{48}$$

uniformly in z. Hence we get from (47)

$$\lim_{n\to\infty}\left\{P[U_n(z)\,|\,\theta_n] - \int_{-\infty}^{z} e^{-\frac{1}{2}(\mu_n^2-2\mu_n t)/c}\,dF_n^*(t)\right\} = 0 \tag{49}$$

uniformly in z. It follows from (49) that for any positive L

$$\lim_{n\to\infty}\left\{P[U_n(L)\,|\,\theta_n] - P[U_n(-L)\,|\,\theta_n] - \int_{-L}^{L} e^{-\frac{1}{2}(\mu_n^2-2\mu_n t)/c}\,dF_n^*(t)\right\} = 0. \tag{50}$$

Since $\lim \mu_n = \mu$, $\lim_{n\to\infty}[F_n^*(t) - F_n(t)] = 0$ uniformly in t, and since $\lim_{n\to\infty} F_n(t) = F(t)$ uniformly in t, we get from (50)

$$\lim_{n\to\infty}\{P[U_n(L)\,|\,\theta_n] - P[U_n(-L)\,|\,\theta_n]\} = \int_{-L}^{L} e^{-\frac{1}{2}(\mu^2-2\mu t)/c}\,dF(t). \tag{51}$$

Now let us calculate the limit of $P[V_n(z)\,|\,\theta_n]$ if $n \to \infty$. The region $V_n(z)$ is defined by the inequality

$$(\hat{\theta}_n - \theta_0)\sqrt{n} < z. \tag{52}$$

This inequality can be written as follows:

$$(\hat{\theta}_n - \theta_n)\sqrt{n} < z - \mu_n\,. \tag{53}$$

Since $\lim \mu_n = \mu$, we get on account of Proposition 1

$$\begin{aligned}\lim_{n\to\infty} P[(\hat{\theta}_n - \theta_n)\sqrt{n} < z - \mu_n\,|\,\theta_n] &= \frac{1}{\sqrt{2\pi c}}\int_{-\infty}^{z-\mu} e^{-\frac{1}{2}t^2/c}\,dt \\ &= \frac{1}{\sqrt{2\pi c}}\int_{-\infty}^{z} e^{-\frac{1}{2}(t-\mu)^2/c}\,dt\end{aligned} \tag{54}$$

Hence

$$\lim_{n\to\infty} P[V_n(z)\,|\,\theta_n] = \frac{1}{\sqrt{2\pi c}}\int_{-\infty}^{z} e^{-\frac{1}{2}(t-\mu)^2/c}\,dt \tag{55}$$

uniformly in z.

For any positive ϵ let L_ϵ denote the positive number satisfying the condition:

$$\frac{1}{\sqrt{2\pi c}}\left[\int_{-\infty}^{-L_\epsilon} e^{-\frac{1}{2}(t-\mu)^2/c}\,dt + \int_{L_\epsilon}^{\infty} e^{-\frac{1}{2}(t-\mu)^2/c}\,dt\right] = \frac{\epsilon}{2}. \tag{56}$$

From (56) we easily get on account of (26)

$$0 \leq \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\mu^2-2\mu t)/c}\, dF(t) - \int_{-L_\epsilon}^{L_\epsilon} e^{-\frac{1}{2}(\mu^2-2\mu t)/c}\, dF(t) \leq \frac{\epsilon}{2}. \tag{57}$$

Since the region $U_n(z_2) - U_n(z_1)$ is a subset of $V_n(z_2) - V_n(z_1)$ for $z_2 > z_1$, we have on account of (55) and (56)

$$\limsup_{n\to\infty} |\{P[U_n(\infty) \mid \theta_n] - P[U_n(L_\epsilon) \mid \theta_n] + P[U_n(-L_\epsilon) \mid \theta_n]\}| \leq \frac{\epsilon}{2}. \tag{58}$$

Since

$$P[U_n(\infty) \mid \theta_n] = G(\mu_n, n),$$

we have

$$\limsup_{n\to\infty} |G(\mu_n, n) - \{P[U_n(L_\epsilon) \mid \theta_n] - P[U_n(-L_\epsilon) \mid \theta_n]\}| \leq \frac{\epsilon}{2}. \tag{59}$$

From (51), (57) and (59) we get

$$\limsup_{n\to\infty} \left| G(\mu_n, n) - \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\mu^2-2\mu t)/c}\, dF(t) \right| \leq \epsilon. \tag{60}$$

Since ϵ can be chosen arbitrarily small, Proposition 2 is proved.

4. Theorems on asymptotically most powerful tests.

THEOREM 1: *Let M_n be the region defined by the inequality $\sqrt{n}\,(\hat{\theta}_n - \theta_0) \geq A_n$, where A_n is chosen such that $P(M_n \mid \theta_0) = \alpha$. Then $\{M_n\}$ is an asymptotically most powerful test of the hypothesis $\theta = \theta_0$, provided the parameter θ is restricted to values $\geq \theta_0$.*

PROOF: Assume that there exists a test $\{W_n\}$ of size α such that

$$\limsup_{n\to\infty} L(W_n, M_n) = \delta > 0. \tag{61}$$

Then there exists a subsequence $\{n'\}$ of the sequence $\{n\}$ and a sequence $\{\theta_{n'}\}$ of parameter values $\geq \theta_0$ such that

$$\lim_{n=\infty} \{P(W_{n'} \mid \theta_{n'}) - P(M_{n'} \mid \theta_{n'})\} = \delta \tag{62}$$

The expression

$$(\theta_{n'} - \theta_0)\sqrt{n} = \mu_{n'} > 0 \tag{63}$$

must be bounded. This can be proved as follows: Since under the assumption $\theta = \theta_0$ the distribution of $\sqrt{n}\,(\hat{\theta}_n - \theta_0)$ converges to a normal distribution with zero mean and finite variance, the sequence $\{A_n\}$ must be bounded. Hence M_n is defined by the inequality

$$\hat{\theta}_n - \theta_0 \geq A_n/\sqrt{n} = \epsilon_n \tag{64}$$

where

$$\lim_{n=\infty} \epsilon_n = 0. \tag{65}$$

From Assumption 1, (64) and (65) it follows easily that if

$$\lim_{n=\infty} \theta_{n'} = \theta_1 > \theta_0, \qquad \lim_{n=\infty} P(M_{n'} \mid \theta_{n'}) = 1.$$

Hence on account of (62) we must have

$$\lim_{n=\infty} \theta_{n'} = \theta_0. \tag{66}$$

If there would exist a subsequence $\{n^*\}$ of $\{n'\}$ such that $\lim_{n=\infty} \mu_{n^*} = \infty$, then on account of (66) and Proposition 1 we would have $\lim_{n=\infty} P(M_{n^*} \mid \theta_{n^*}) = 1$, which is in contradiction to (62). Hence the expression (63) must be bounded. Let $\{n''\}$ be a subsequence of $\{n'\}$ such that

$$\lim_{n=\infty} \mu_{n''} = \mu > 0. \tag{67}$$

Denote by $F_n(z)$ the probability of the intersection of W_n and the region $(\hat{\theta}_n - \theta_0)\sqrt{n} < z$ under the hypothesis that $\theta = \theta_0$. Consider the subsequence $\{n'''\}$ of the sequence $\{n''\}$ such that $F_{n'''}(z)$ converges with $n \to \infty$ towards a function $F(z)$. The existence of such a subsequence $\{n'''\}$ can be proved as follows: Denote the probability $P[(\hat{\theta}_n - \theta_0)\sqrt{n} < z \mid \theta_0]$ by $\Phi_n(z)$. On account of Proposition 1, $\Phi_n(z)$ converges with $n \to \infty$ uniformly in z towards

$$\psi(z) = \frac{1}{\sqrt{2\pi c}} \int_{-\infty}^{z} e^{-\frac{1}{2c}t^2} dt \tag{68}$$

where c has the same value in (23).
We obviously have

$$F_n(z_2) - F_n(z_1) \leq \Phi_n(z_2) - \Phi_n(z_1) \tag{69}$$

for any pair of values z_1, z_2 for which $z_2 > z_1$. Hence

$$\limsup_{n=\infty} [F_n(z_2) - F_n(z_1)] \leq \psi(z_2) - \psi(z_1). \tag{70}$$

Since $F_n(z)$ is a monotonic function of z, our statement follows easily from (70) and the fact that $\psi(z)$ is uniformly continuous. Hence on account of Proposition 2 we have

$$\lim_{n=\infty} P(W_{n'''} \mid \theta_{n'''}) = \int_{-\infty}^{\infty} e^{-\frac{1}{2c}(\mu^2 - 2\mu z)} dF(z) \tag{71}$$

and

$$\lim_{n=\infty} P(M_{n'''} \mid \theta_{n'''}) = \int_{-\infty}^{\infty} e^{-\frac{1}{2c}(\mu^2 - 2\mu z)} d\Phi(z) \tag{72}$$

where

$$\Phi(z) = 0 \text{ for } z \leq z_0, \tag{73}$$

$$\Phi(z) = \psi(z) - \psi(z_0) \text{ for } z > z_0, \tag{74}$$

and z_0 is given by

$$1 - \psi(z_0) = \alpha. \tag{75}$$

From (62), (71) and (72) we get

$$\int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - 2\mu v)} d[F(z) - \Phi(z)] = \delta > 0. \tag{76}$$

Consider a normally distributed variate y with mean ν and variance c. Let B be a critical region of size α for testing the hypothesis $\nu = 0$ by a single observation on y, i.e. B is a subset of the real axis $[-\infty, +\infty]$. Denote by $D(v)$ the intersection of B and the region $C(v)$ defined by the inequality $y < v$. Denote by $H(v)$ the probability of $D(v)$ under the hypothesis $\nu = 0$. Then the power of the test B with respect to the alternative $\nu = \mu$ is given by the following expression

$$\int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - \mu v)} dH(v). \tag{77}$$

If the region B is given by the inequality $y \geq v_0$ where v_0 is chosen such that the size of B is equal to α, then $H(v) = \Phi(v)$ where the function Φ is defined by the equations (73), (74) and (75). Since the latter test is uniformly most powerful[4] with respect to all alternatives $\nu > 0$, for any positive μ the inequality

$$\int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - \mu v)} d[H(v) - \Phi(v)] \leqq 0 \tag{78}$$

holds. Let

$$\psi(v) = \frac{1}{\sqrt{2\pi c}} \int_{-\infty}^{v} e^{-\frac{1}{2}t^2/c} dt.$$

It is obvious that

$$H(v_2) - H(v_1) \leq \psi(v_2) - \psi(v_1) \text{ for } v_2 > v_1 \tag{79}$$

and

$$\int_{-\infty}^{\infty} dH(v) = \alpha. \tag{80}$$

---

[4] See for instance J Neyman and E S. Pearson, "Contributions to the theory of testing statistical hypotheses," *Stat Res Memoirs*, Vol. 1 (1936)

On the other hand, if $K(v)$ is a monotonically non-decreasing non-negative function of v such that

$$K(v_2) - K(v_1) \leq \psi(v_2) - \psi(v_1) \text{ for } v_2 > v_1 \tag{79'}$$

and

$$\int_{-\infty}^{\infty} dK(v) = \alpha \tag{80'}$$

hold, then there exists a sequence $\{B^{(i)}\}$, $(i = 1, 2, \cdots, \text{ad inf.})$, of regions of size α such that

$$\lim_{i=\infty} H^{(i)}(v) = K(v)$$

uniformly in v. Since (78) holds for $H(v) = H^{(i)}(v)$, and since

$$H^{(i)}(v_2) - H^{(i)}(v_1) \leq \psi(v_2) - \psi(v_1) \text{ for } v_2 > v_1,$$

it is easy to see that (78) will hold also for $H(v) = K(v)$. Hence for any monotonically non-decreasing non-negative function $K(v)$ for which (79') and (80') are fulfilled, also (78) must hold. Since $F(v)$ is a distribution function which satisfies (79') and (80'), we have a contradiction to (76). This proves Theorem 1.

THEOREM 2: *Let M_n be the region defined by the inequality $\sqrt{n}\,(\hat{\theta}_n - \theta_0) \leq A_n$, where A_n is chosen such that $P(M_n \mid \theta_0) = \alpha$. Then $\{M_n\}$ is an asymptotically most powerful test of the hypothesis $\theta = \theta_0$, provided that the parameter θ is restricted to values $\leq \theta_0$.*

We omit the proof since it is entirely analogous to that of Theorem 1.

THEOREM 3: *Let M_n be the region consisting of all points which satisfy at least one of the inequalities*

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \leq -A_n, \qquad \sqrt{n}\,(\hat{\theta}_n - \theta_0) \geq A_n.$$

The constant $A_n > 0$ is chosen such that $P(M_n \mid \theta_0) = \alpha$. Then $\{M_n\}$ is an asymptotically most powerful unbiased test of the hypothesis $\theta = \theta_0$.

PROOF: Assume that there exists a sequence $\{W_n\}$ $(n = 1, 2, \cdots, \text{ad inf.})$ of regions such that

$$P(W_n \mid \theta_0) = \alpha \tag{81}$$

$$\lim_{n=\infty} g(W_n) = \alpha \tag{82}$$

and

$$\limsup_{n=\infty} L(W_n, M_n) = \delta > 0. \tag{83}$$

We shall deduce a contradiction from this assumption. On account of (83) there exists a subsequence $\{n'\}$ of $\{n\}$ such that

$$\lim_{n=\infty} \{P(W_{n'} \mid \theta_{n'}) - P(M_{n'} \mid \theta_{n'})\} = \delta. \tag{84}$$

The expression

$$(\theta_{n'} - \theta_0)\sqrt{n'} = \mu_{n'} \tag{85}$$

must be bounded. The proof of this statement is omitted, since it is analogous to the proof of the similar statement about (63). Hence there exists a subsequence $\{n''\}$ of $\{n'\}$ such that

$$\lim_{n=\infty} \mu_{n''} = \mu. \tag{86}$$

Denote by $F_n(z)$ the probability of the intersection of W_n with the region $(\hat{\theta}_n - \theta_0)\sqrt{n} < z$ under the hypothesis $\theta = \theta_0$. Consider a subsequence $\{n'''\}$ of $\{n''\}$ such that $F_{n'''}(z)$ converges with $n \to \infty$ towards a function $F(z)$. The existence of such a sequence $\{n'''\}$ can be proved in the same way as the similar statement in the proof of Theorem 1. Hence on account of Proposition 2 and (86) we have

$$\lim_{n=\infty} P(W_{n'''} \mid \theta_{n'''}) = \int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - 2\mu z)/c}\, dF(z) \tag{87}$$

and

$$\lim_{n=\infty} P(M_{n'''} \mid \theta_{n'''}) = \int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - 2\mu z)/c}\, d\Phi(z) \tag{88}$$

where

$$\Phi(z) = \frac{1}{\sqrt{2\pi c}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2/c}\, dt \quad \text{for} \quad z \leq -z_0, \tag{89}$$

$$\Phi(z) = \Phi(-z_0) \quad \text{for} \quad -z_0 \leq z \leq z_0 \tag{90}$$

$$\Phi(z) = \Phi(-z_0) + \frac{1}{\sqrt{2\pi c}} \int_{z_0}^{z} e^{-\frac{1}{2}t^2/c}\, dt \quad \text{for} \quad z > z_0, \tag{91}$$

and

$$\Phi(-z_0) = \tfrac{1}{2}\alpha. \tag{92}$$

From (84), (87) and (88) it follows that

$$\int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - 2\mu z)/c}\, d[F(z) - \Phi(z)] = \delta. \tag{93}$$

Consider a normally distributed variate y with means ν and variance c. Let B an unbiased critical region of size α for testing the hypothesis $\nu = 0$ by a single observation on y, i.e. B is a subset of the real axis $[-\infty, +\infty]$. Denote by $D(v)$ the intersection of B with the region $C(v)$ defined by the inequality $y < v$. Denote by $H(v)$ the probability of $D(v)$ under the hypothesis $\nu = 0$. Then the power of the test B with respect to the alternative $\nu = \mu$ is given by

$$\int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - 2\mu v)/c}\, dH(v). \tag{94}$$

If the region B consists of all points which satisfy at least one of the inequalities $y \leq -v_0$, $y \geq v_0$, and if $v_0 > 0$ is chosen such that the size of B is equal to α, then $H(v) = \Phi(v)$, where $\Phi(v)$ is defined by the equations (89)–(92). Since the latter test is a uniformly most powerful unbiased test,[5] for any μ the inequality

$$\int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - 2\mu v)/c}\, d[H(v) - \Phi(v)] \leq 0 \tag{95}$$

holds. Let

$$\psi(v) = \frac{1}{\sqrt{2\pi c}} \int_{-\infty}^{v} e^{-\frac{1}{2}t^2/c}\, dt.$$

It is obvious that

$$H(v_2) - H(v_1) \leq \psi(v_2) - \psi(v_1) \quad \text{for} \quad v_2 > v_1, \tag{96}$$

$$\int_{-\infty}^{\infty} dH(v) = \alpha \tag{97}$$

and

$$\int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - 2\mu v)/c}\, dH(v) \text{ has a minimum for } \mu = 0, \tag{98}$$

On the other hand, if $K(v)$ is a monotonically non-decreasing non-negative function of v such that

$$K(v_2) - K(v_1) \leq \psi(v_2) - \psi(v_1) \text{ for } v_2 > v_1, \tag{96'}$$

$$\int_{-\infty}^{\infty} dK(v) = \alpha, \tag{97'}$$

$$\int_{-\infty}^{\infty} e^{-\frac{1}{c}(\mu^2 - 2\mu v)/c}\, dK(v) \text{ has a minimum for } \mu = 0, \tag{98'}$$

then there exists a sequence $\{B^{(i)}\}$ ($i = 1, 2, \cdots$, ad inf.) of unbiased regions of size α such that

$$\lim_{i=\infty} H^{(i)}(v) = K(v)$$

uniformly in v. Since (95) holds for $H(v) = H^{(i)}(v)$ ($i = 1, 2, \cdots$, ad inf.), and since

$$H^{(i)}(v_2) - H^{(i)}(v_1) \leq \psi(v_2) - \psi(v_1) \text{ for } v_2 > v_1,$$

it is easy to see that (95) holds also for $H(v) = K(v)$. Hence for any monotonically non-decreasing non-negative function $K(v)$ for which (96'), (97'), and (98') are fulfilled, also (95) must be fulfilled if we substitute $K(v)$ for $H(v)$.

[5] J. Neyman and E. S. Pearson, l. c., p. 29.

Since $F(v)$ is a distribution function which satisfies (96'), (97') and (98'), we have a contradiction to (93). This proves Theorem 3.

5. Appendix. *Proof of the uniform consistency of* $\hat{\theta}_n$. It will be shown here that under certain conditions on the density function $f(x, \theta)$, Assumption 1, i.e. uniform consistency of $\hat{\theta}_n$, can be proved.

For any open subset ω of the θ-axis we denote by $\varphi(x, \omega)$ the least upper bound, and by $\psi(x, \omega)$ the greatest lower bound of $\frac{\partial^2 \log f(x, \theta)}{\partial \theta^2}$ with respect to θ in the set ω. For any function $\lambda(x)$ we denote by $E_\theta \lambda(x)$ the expected value of $\lambda(x)$ under the assumption that θ is the true value of the parameter, i.e.

$$E_\theta \lambda(x) = \int_{-\infty}^{\infty} \lambda(x) f(x, \theta)\, dx.$$

Denote furthermore by $P(\hat{\theta}_n \in \omega \mid \theta)$ the probability that $\hat{\theta}_n$ will fall in ω under the assumption that θ is the true value of the parameter. Finally denote by Ω the parameter space and assume that Ω is either the whole real axis or a subset of it.

PROPOSITION 3. *$\hat{\theta}_n$ is a uniformly consistent estimate of θ, i.e. for any positive k*

$$\lim_{n=\infty} P(-k < \hat{\theta}_n - \theta < k \mid \theta) = 1$$

uniformly for all θ in Ω, if the following two conditions are fulfilled:

Condition I. For all values θ in Ω

$$\int_{-\infty}^{\infty} \frac{\partial f(x, \theta)}{\partial \theta}\, dx = \int_{-\infty}^{\infty} \frac{\partial^2 f(x, \theta)}{\partial \theta^2}\, dx = 0.$$

Condition II. For any value θ in Ω there exists an open interval $\omega(\theta)$ containing θ and having the following three properties:

II_a. $$\lim_{n=\infty} P(\hat{\theta}_n \in \omega(\theta) \mid \theta] = 1$$

uniformly for all θ in Ω.

II_b. *$E_\theta \varphi^2[x, \omega(\theta)]$ is a bounded function of θ in Ω, and the least upper bound A of $E_\theta \varphi[x, \omega(\theta)]$ with respect to θ in Ω is negative.*

II_c. *$E_\theta \psi[x, \omega(\theta)]$ is a bounded function of θ in the set Ω.*

Condition I means simply that we may differentiate under the integral sign. In fact

$$\int_{-\infty}^{\infty} f(x, \theta) = 1$$

identically in θ. Hence

$$\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(x, \theta)\, dx = \frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{\infty} f(x, \theta)\, dx = 0.$$

Differentiating under the integral sign, we obtain Condition I.

In case that $\omega(\theta)$ is the whole axis Condition II_a reduces to the condition that $\hat{\theta}_n$ exists.

In order to prove Proposition 3, we show first that for any positive η

$$\lim_{n=\infty} P\left[\left(-\eta < \frac{1}{n}\sum_{\alpha=1}^{n}\frac{\partial \log f(x_\alpha, \theta)}{\partial\theta} < \eta\right)\Big|\,\theta\right] = 1 \tag{99}$$

uniformly for all θ in Ω. We have on account of Condition I

$$E_\theta \frac{\partial \log f(x, \theta)}{\partial\theta} = E_\theta \frac{\partial f(x, \theta)}{\partial\theta}\Big/ f(x, \theta) = \int_{-\infty}^{\infty}\frac{\partial f(x, \theta)}{\partial\theta}\,dx = 0. \tag{100}$$

Since

$$\frac{\partial^2 \log f(x, \theta)}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left[\frac{\partial f(x, \theta)}{\partial\theta}\Big/ f(x, \theta)\right] = \frac{\partial^2 f(x, \theta)}{\partial\theta^2}\Big/ f(x, \theta) - \left\{\frac{\partial f(x, \theta)}{\partial\theta}\Big/ [f(x, \theta]^2\right\}^2$$

we have on account of Condition I

$$E_\theta\left(\frac{\partial \log f(x, \theta)}{\partial\theta}\right)^2 = -E_\theta\frac{\partial^2 \log f(x, \theta)}{\partial\theta^2}. \tag{101}$$

According to Condition II $E_\theta\psi[x, \omega(\theta)] < 0$ and is a bounded function of θ. Since $E_\theta\dfrac{\partial^2 \log f(x, \theta)}{\partial\theta^2} < 0$ and $> E_\theta\psi[x, \omega(\theta)]$, the left hand side of (101), i.e. the variance of $\dfrac{\partial \log f(x, \theta)}{\partial\theta}$, is a bounded function of θ. From this and the equation (100) we obtain easily (99). Consider the Taylor expansion

$$\frac{1}{n}\sum_\alpha \frac{\partial \log f(x_\alpha, \theta)}{\partial\theta} = (\theta - \hat{\theta}_n)\frac{1}{n}\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \theta'_n)}{\partial\theta^2}, \tag{102}$$

where θ'_n lies in the interval $[\theta, \hat{\theta}_n]$. Let ϵ be an arbitrary positive number and denote by $Q_n(\theta)$ the region defined by the inequality

$$\left|\frac{1}{n}\sum_\alpha \frac{\partial \log f(x_\alpha, \theta)}{\partial\theta}\right| \leq \epsilon. \tag{103}$$

On account of (99) we have

$$\lim_{n=\infty} P[Q_n(\theta) \mid \theta] = 1 \tag{104}$$

uniformly for all θ in Ω.

Denote by $R_n(\theta)$ the region defined by the inequality

$$\frac{1}{n}\sum_\alpha \varphi[x_\alpha, \omega(\theta)] < \tfrac{1}{2}A < 0. \tag{105}$$

On account of Condition II_b

$$\lim_{n=\infty} P[R_n(\theta) \mid \theta] = 1 \tag{106}$$

uniformly for all θ in Ω. Denote by $B_n(\theta)$ the region in which $\hat{\theta}_n \in \omega(\theta)$. Since in $B_n(\theta)$

$$\frac{1}{n}\sum \frac{\partial^2 \log f(x_\alpha, \theta'_\alpha)}{\partial\theta^2} \leq \frac{1}{n}\sum \varphi[x_\alpha, \omega(\theta)]$$

we have in the intersection $R'_n(\theta)$ of $R_n(\theta)$ and $B_n(\theta)$

$$\left|\frac{1}{n}\sum \frac{\partial^2 \log f(x_\alpha, \theta'_\alpha)}{\partial\theta^2}\right| > \left|\frac{A}{2}\right|. \tag{107}$$

Denote by $U_n(\theta)$ the intersection of $Q_n(\theta)$ and $R'_n(\theta)$. It is obvious that

$$\lim_{n=\infty} P[U_n(\theta) \mid \theta] = 1 \tag{108}$$

uniformly for all θ in Ω. From (102), (103) and (107) we get that in $U_n(\theta)$

$$|\theta - \hat{\theta}_n| \leq \frac{\epsilon}{\left|\frac{1}{2}A\right|} = \frac{2\epsilon}{|A|}. \tag{109}$$

Hence on account of (108)

$$\lim_{n=\infty} P\left(|\theta - \hat{\theta}_n| < \frac{2\epsilon}{|A|}\,\middle|\,\theta\right) = 1$$

uniformly for all θ in Ω. Since ϵ can be chosen arbitrarily, Proposition 3 is proved.

Conditions I and II are sufficient but not necessary for the uniform consistency of $\hat{\theta}_n$. For sufficiently small $\omega(\theta)$ the conditions II_b and II_c are rather weak. In fact, on account of (101) we have

$$E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^2} < 0.$$

Hence for sufficiently small intervals $\omega(\theta)$, under certain continuity conditions, also $E_\theta\varphi[x, \omega(\theta)]$ will be negative. However, in some cases may be difficult to verify II_a for small $\omega(\theta)$. On the other hand, for sufficiently large $\omega(\theta)$ (certainly for $\omega(\theta) = [-\infty, +\infty]$) II_a can easily be verified, but the conditions II_b and II_c might be unnecessarily strong. In cases where II_b or II_c does not hold for $\omega(\theta) = [-\infty, +\infty]$ and the validity of II is not apparent, the following Lemma may be useful:

LEMMA: *Proposition* 3 *remains valid if we substitute for Condition II the conditions*

II'. *Denote by* T_n *the set of all points at which* $\hat{\theta}_n$ *exists and*

$$\sum_\alpha \frac{\partial}{\partial\theta} \log f(x_\alpha, \theta^*) = 0 \tag{110}$$

has at most one solution in θ^*. *Then* $\lim_{n=\infty} P[T_n \mid \theta] = 1$ *uniformly for all* θ *in* Ω, *and*

II''. *There exists a positive* k *such that for* $\omega(\theta) = I(\theta) = (\theta - k, \theta + k)$ *the following two conditions hold:*

II_b''. *$E_\theta\varphi^2[x, I(\theta)]$ is a bounded function of θ in Ω and the least upper bound A of $E_\theta\varphi[x, I(\theta)]$ with respect to θ in Ω is negative.*

II_c''. *$E_\theta\psi[x, I(\theta)]$ is a bounded function of θ in the set Ω.* In cases where II_b or II_c is not fulfilled for $\omega(\theta) = [-\infty, +\infty]$ the verification of II and II'' may be easier than that of II.

Our Lemma can be proved as follows: Consider the Taylor expansion

$$\frac{1}{n}\Sigma\frac{\partial}{\partial\theta}\log f(x_\alpha, \theta^*) = \frac{1}{n}\Sigma\frac{\partial}{\partial\theta}\log f(x_\alpha, \theta) + (\theta^* - \theta)\frac{1}{n}\Sigma\frac{\partial^2}{\partial\theta^2}\log f(x_\alpha, \theta') \tag{111}$$

where θ' lies in $[\theta, \theta^*]$. Denote by $V_n(\theta)$ the region defined by

$$\frac{1}{n}\Sigma\varphi[x_\alpha, I(\theta)] < \tfrac{1}{2}A < 0. \tag{112}$$

On account of II_b'' we have

$$\lim_{n=\infty} P[V_n(\theta) \mid \theta] = 1 \tag{113}$$

uniformly for all θ in Ω. Let $W_n(\theta)$ be the region defined by

$$\left|\frac{1}{n}\Sigma\frac{\partial}{\partial\theta}\log f(x_\alpha, \theta)\right| < \epsilon \tag{114}$$

From Condition I and Condition II_c'' it follows easily that

$$\lim_{n=\infty} P[W_n(\theta) \mid \theta] = 1 \tag{115}$$

uniformly for all θ in Ω. For all values θ^* in the interval $I(\theta)$ we have

$$\frac{1}{n}\Sigma\varphi[x_\alpha, I(\theta)] \geq \frac{1}{n}\Sigma\frac{\partial^2}{\partial\theta^2}\log f(x_\alpha, \theta'). \tag{116}$$

Because of (112) and (116) we have in $V_n(\theta)$

$$\frac{1}{n}\Sigma\frac{\partial^2}{\partial\theta^2}\log f(x_\alpha, \theta') < \tfrac{1}{2}A < 0 \tag{117}$$

for all values θ^* in the interval $I(\theta)$ Let ϵ be less than $|\frac{1}{4}kA|$. Then in the intersection $W_n'(\theta)$ of the regions $V_n(\theta)$ and $W_n(\theta)$ we obviously have on account of (114) that the values of the left hand side of (111) for $\theta^* = \theta + k$ and $\theta^* = \theta - k$ will be of opposite sign. Hence at any point of $W_n'(\theta)$ the equation (110) has at least one root which lies in the interval $I(\theta)$. Since (110) has at most one root in T_n and since $\hat{\theta}_n$ is a root of (110), we get that at any point of the intersection $W_n''(\theta)$ of $W_n'(\theta)$ and T_n, $\hat{\theta}_n$ lies in $I(\theta)$. Since

$$\lim_{n=\infty} P[W''(\theta) \mid \theta] = 1 \quad \text{uniformly for all } \theta \text{ in } \Omega, \tag{118}$$

also

$$\lim_{n=\infty} P[\hat{\theta}_n \epsilon I(\theta) \mid \theta] = 1 \tag{119}$$

uniformly for all θ in Ω. The relation (119) combined with the conditions II_b'' and II_c'' is equivalent to Condition II. Hence our Lemma is proved.

EXPERIMENTAL DETERMINATION OF THE MAXIMUM OF A FUNCTION[1]

By Harold Hotelling

Columbia University, New York City

1. The necessary background for efficient experimental determinations. We shall deal with the problem of arranging an experiment for determining the value of x for which an unknown function $f(x)$ is a maximum or minimum. This problem is to be distinguished from those of estimating the maximum or minimum itself, and of studying the distributions of such estimates, problems to which Bernstein [1] and Rice [2] have contributed.

The range of applications in which determinations of maximizing and minimizing values are important is extremely wide. Among these are the determination of the time of year at which the number of algae or bacilli in a lake is a maximum, and the amount of fertilizers and of irrigation water making the yield of a crop a maximum. The magnetic permeabilities of permalloys, perminvars and permendurs as functions of the induction, and the hardness of a copper-iron alloy as a function of the time of aging at 500°C., possess smooth maxima having interest in telephony, [3], [4]. The effective range of a gun is a function of the speed of burning of the powder, a variable which can be controlled. Almost every entrepreneur has a fervent desire to know the selling prices that will yield a maximum profit, and a few have undertaken controlled experiments with a view to finding out. There are also numerous practical problems of minimizing costs; for example, the cost of operating a ship as a function of its speed possesses a minimum. We shall confine our attention chiefly to the experimental determination of maxima, since such problems seem to occur naturally with greater frequency in applications; there is no loss of generality in this, since $f(x)$ has a maximum where $-f(x)$ has a minimum.

We shall assume that, for each value of x in the set we shall select, one or more observations will be made on $y = f(x)$, and that these observations are afflicted with errors which are independently distributed about zero with a common variance σ^2. From this it follows that if $f(x)$ is a linear function of known functions of x, with unknown coefficients $\beta_0, \beta_1, \cdots, \beta_p$ (for example a polynomial in x), the most efficient method of fitting is the method of least squares, which yields unbiased estimates $b_0, \cdots, b_p$ of $\beta_0, \cdots, \beta_p$ having the least possible variances; this is true whether or not the errors are normally distributed. If the fourth moment of the errors is finite, and if the number N

[1] Presented at the joint meeting of the Institute of Mathematical Statistics and the American Mathematical Society at Hanover, September 10, 1940.

of observations is large, the estimated coefficients will be distributed in an approximately normal manner; and so also will any function of them that is regular in a fixed neighborhood of its "population value." By the "population value" of a function $\phi(b_0, \cdots, b_p)$ we mean $\phi(\beta_0, \cdots, \beta_p)$. In particular, if

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 \cdots \beta_p x^p$$

has a maximum for $x = \xi$ of the simplest type, such that $f'(\xi) = 0$ and $f''(\xi) < 0$, so that ξ is a simple root of the equation

$$f'(\xi) = \beta_1 + 2\beta_2 \xi + \cdots + p\beta_p \xi^{p-1} = 0,$$

and if x_0 is an estimate of ξ found from the polynomial fitted by the method of least squares, so that

$$b_1 + 2b_2 x_0 + \cdots + pb_p x_0^{p-1} = 0,$$

this last equation defines x_0 as a function of $b_1, \cdots, b_p$. The function is, to be sure, multiple-valued when $p > 2$; but for sufficiently large values of N the probability will become arbitrarily great that the roots obtained from a random experiment will each differ by an arbitrarily small quantity from one of the roots of $f'(x) = 0$. Then *provided we have a sufficient preliminary approximate knowledge of* ξ, we may choose the root nearest ξ; and the probability distribution of this root, which in nearly all experiments will be a single-valued function

$$\phi(b_1, \cdots, b_p),$$

will approach normality of form, with standard error of order $N^{-1/2}$, about a mean differing from

$$\xi = \phi(\beta_1, \cdots, \beta_p)$$

at most by terms of order N^{-1}, which are thus negligible in comparison with the standard error. The situation will be effectively the same if, without knowing ξ in advance even approximately, we choose the root x_0 giving the greatest value $f(x_0)$, provided $f(\xi)$ is greater than any other value of $f(x)$.

From these considerations it appears advisable, whenever the unknown function is capable of being represented adequately by a polynomial of degree p considerably less than the number N of observations, to fit a polynomial of degree p by least squares, and from it to determine the maximizing value by differentiation. In practice, however, there are obstacles to carrying out such a procedure with confidence. The form of the function is usually not known; it is far from clear what value should be given p even if the function is to be regarded as a polynomial; the use of a polynomial which does not give a sufficiently good fit, with observations taken at a considerable distance from the maximizing value, perhaps separated from it by other maxima and minima, appears to be a highly dubious proceeding; and if p is taken large, the labor of calculation becomes excessive. For all these reasons it is desirable to assign the values of x which are to be the basis of the experimental work close enough to

the maximizing value ξ so that a polynomial of very low degree will fit adequately in the neighborhood.

We shall restrict ourselves to functions having continuous derivatives of all relevant orders[2] in a neighborhood of ξ. Such a function can in a sufficiently small neighborhood be approximated by a polynomial of the second degree. The necessity of using a polynomial of higher degree can therefore be avoided, *when a fairly good knowledge of the function is already in hand*, and when the number N of observations that can be made is large enough, by choosing all the values of x in a sufficiently small neighborhood of ξ. We shall suppose that this is done; that is, a regression equation

$$Y = b_0 + b_1 x + b_2 x^2$$

is fitted by least squares to a large number of observations after choosing the values of x quite close to the true maximizing value ξ; and the estimate x_0 of ξ is a solution of $dY/dx = b_1 + 2b_2 x = 0$, so that

$$x_0 = -\frac{b_1}{2b_2}.$$

We shall examine the errors in x_0 arising both from the inadequacy that may exist in the quadratic approximation and from the random errors of observation, and shall consider what distribution of x may most appropriately be chosen to reduce the errors of both kinds, and to place them in a suitable balance with each other.

It will be observed that a fairly definite preliminary knowledge of the function under investigation is required for such a program. Any criterion for the selection of values of x for experimentation must involve not only the value of ξ but also the values of the first few derivatives in a neighborhood of ξ, or some similar information. The requirement of preliminary information is essential for the efficient design of experiments in general. For instance the efficiency of an agricultural field experiment depends on the correctness of the appraisal, before the experiment is laid down, of the general nature of the fertility gradients likely to exist in the field and of the variances due to error and main effects which will be revealed more accurately by the experiment itself. If the pre-

[2] Other cases may well arise in practice and deserve separate consideration in connection with the particular investigations in which they arise. For example various physical properties of alloys, regarded as functions of the proportion of a particular constituent, have maxima, but may have discontinuous derivatives because of the phenomena of crystallization and solution of one metal in another. The assumptions appropriate to an investigation, parallel to that of the present paper, of the proper organization of experiments for finding such metallurgical maxima must be drawn from metallurgy. The case of continuous derivatives is however of widespread importance. If *no* regularity assumption is made about the function, one set of N values of x is as good as another, and no set is likely to tell us very much about the function if it is one of the violently irregular ones utilized in the theory of functions to emphasize the necessity of studying that subject.

liminary information is incorrect, a properly arranged self-contained experiment will nevertheless give results which are *valid*, in the sense that the significance probabilities calculated from them by accurate methods are correct, but will be *inefficient*, in the sense that another experiment of the same cost, based on better preliminary information, would be more likely to detect real effects through the smallness of such a calculated probability. The efficient conduct of experimentation thus proceeds in stages of ascending magnitude. A large-scale investigation should be preceded by a smaller one designed primarily to obtain information for use in designing the large one. The small preliminary investigation may well in turn be preceded by a still smaller pre-preliminary investigation, and so on,[3] like an army marching after an advance guard, which follows a more advanced smaller detachment, which follows a still smaller and still more advanced unit, which follows a "point." At the very beginning of the process of chain experimentation will stand work based on little or no clear information of the kind required for efficient design. This first phase will be speculative and exploratory in character. Neither its cost nor its accuracy can well be estimated in advance. It is a favorite, but not exclusive, preoccupation of men of genius. Many of its results turn out to be worthless. But it is an essential preliminary to well-organized research directed to definite aims defined qualitatively in advance.

After the first speculative and unsystematic phase in the knowledge of a subject is past, but before the careful, economical organization of an accurate investigation, an intermediate type of exploration is needed to supply estimates of the parameters required for the design of the full-scale investigation. In the present case such a systematic though small-scale experiment might perhaps consist in dividing a range within which the desired maximizing value ξ is known to lie into equal parts, making at least two observations at each of the ends of these intervals, and fitting a polynomial of at least the fifth degree by least squares. This will make possible estimates of the parameters $\sigma, \beta_1, \beta_2, \cdots, \beta_5$ (and hence of ξ) required for using the efficient designs which we shall obtain. At least six different values of x are required for fitting the polynomial of the fifth degree. The fitting process is facilitated by taking them in arithmetic progression and using orthogonal polynomials.

---

[3] A remarkable example of such a series of investigations is the chain of sample censuses of area of jute in Bengal carried out for the Indian Central Jute Committee under the direction of Prof. P. C. Mahalanobis annually beginning in 1937. Each year's work is designed primarily to obtain information for planning the next year's, and a sequence of four or five such investigations, each considerably larger than the preceding, is planned to lead up to an eventual annual sampling of the whole immense jute area in the province. A partial account of this is given in [5], a fuller one in confidential but printed reports of the Indian Central Jute Committee, Calcutta.

Certain multiple-sample schemes in manufacturing inspection also provide good examples of chain experiments, [6].

2. Sampling errors and bias in the quadratic approximation. Let us measure all values of x from the value ξ under investigation which makes $f(x)$ a maximum. Then $\xi = 0$, and in the expansion

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots \tag{1}$$

we shall have $\beta_1 = 0$ and $\beta_2 \leq 0$; we shall assume that $\beta_2 < 0$. An observation y_α corresponding to a chosen value x_α will have, by assumption, an error Δ_α of zero expectation and variance σ^2, such that

$$y_\alpha = f(x_\alpha) + \Delta_\alpha . \tag{2}$$

A quadratic estimate

$$Y = b_0 + b_1 x + b_2 x^2 \tag{3}$$

of $f(x)$ is obtained by means of normal equations which may be written

$$\begin{aligned} a_0 b_0 + a_1 b_1 + a_2 b_2 &= Sy \\ a_1 b_0 + a_2 b_1 + a_3 b_2 &= Sxy \\ a_2 b_0 + a_3 b_1 + a_4 b_2 &= Sx^2 y, \end{aligned} \tag{4}$$

where S stands for summation over all the observations, so that, for example, $Sy = \Sigma y_\alpha = y_1 + y_2 + \cdots + y_N$, and where

$$a_k = Sx^k. \tag{5}$$

In particular, $a_0 = N$. A determinate solution is possible only if there are at least three distinct values of x; we shall always suppose therefore that this is the case. This is equivalent to assuming that the determinant a of the coefficients in (4) is not zero. A greater number of observations y is necessary to obtain an estimate of the variance σ^2, and furthermore we shall suppose this number large in our approximations, but since repeated observations may be made for each value of x, it is not essential that there be more than three values of x in the distribution to be selected.

If we put

$$\delta b_k = b_k - \beta_k , \qquad \gamma_k = Sx^k \Delta, \tag{6}$$

for $k = 0, 1, 2$, substitute (1) in (2) and the result in (4), and utilize (5) and (6), we obtain

$$\begin{aligned} a_0 \delta b_0 + a_1 \delta b_1 + a_2 \delta b_2 &= \gamma_0 + a_3 \beta_3 + a_4 \beta_4 + \cdots \\ a_1 \delta b_0 + a_2 \delta b_1 + a_3 \delta b_2 &= \gamma_1 + a_4 \beta_3 + a_5 \beta_4 + \cdots \\ a_2 \delta b_0 + a_3 \delta b_1 + a_4 \delta b_2 &= \gamma_2 + a_5 \beta_3 + a_6 \beta_4 + \cdots \end{aligned} \tag{7}$$

From these equations it follows that the errors δb_k are homogeneous linear functions of the right-hand members and will therefore be small if the quantities on the right are small. Of these quantities, the γ_k's will be stochastically of the

order $N^{1/2}$ for large samples with any fixed set of values of x. When the equations are solved, their coefficients will be of the order of N^{-1}, so that the product is of order $N^{-1/2}$, and becomes negligible if N is large enough. The coefficients a_k of $\beta_3, \beta_4, \cdots$ can be kept small if the values of x are chosen to lie within a sufficiently restricted range. Of course the coefficients a_k in the *left* members of (7) will also be small in this case, but not small enough to offset fully the smallness of those on the right. To see this, we observe that if all the values of x be multiplied by any quantity g, a_k is multiplied by g^k, while

$$a = \begin{vmatrix} a_0 & a_1 & a_2 \\ a_1 & a_2 & a_3 \\ a_2 & a_3 & a_4 \end{vmatrix} \tag{8}$$

is multiplied by g^6. The cofactors of the last column are proportional respectively to g^4, g^3 and g^2. Hence, in the expression for δb_2, the coefficient of β_3 is of order g, that of β_4 is of order g^2, and so on, the coefficients of the β's of higher orders vanishing more and more rapidly with g as we go on in the sequence. The like is true of δb_1 and δb_0, which vanish even more rapidly with g. Thus we may, by restricting sufficiently the range of x on the basis of the assumed preliminary knowledge of the function, and taking a sufficiently large sample of observations, bring it about that the probability will be arbitrarily close to unity that the δb_k's are less than any assigned limits.

Let us, in particular, restrict the range sufficiently and take a large enough sample to make it reasonable to regard δb_2 as negligible in comparison with β_2. The error in the estimate

$$x_0 = -\frac{b_1}{2b_2} \tag{9}$$

of the maximizing value ξ will, since we are taking $\xi = 0$, be x_0 itself, and may be written

$$\delta x_0 = -\frac{\delta b_1}{2(\beta_2 + \delta b_2)} = -\tfrac{1}{2}\frac{\delta b_1}{\beta_2}\left(1 - \frac{\delta b_2}{\beta_2} + \cdots\right),$$

where the terms other than 1 in the last parentheses are negligible. The problem of minimizing the error δx_0 is then virtually equivalent to minimizing the error δb_1. In section 5 it will be shown that it is not until we reach terms of the order of g^5 that the errors δb_2 need be taken into account. We shall first discuss the errors in x_0 of lower orders in g, and thus confine the discussion to δb_1. For the present we shall take as the quantity to be made as small as possible the expectation of the square of this last error, $E(\delta b_1)^2$. This is not the same as the variance of b_1, since $E\delta b_1$ is not in general zero. We have, in fact, by transposing a familiar formula for the variance,

$$E(\delta b_1)^2 = (E\delta b_1)^2 + \sigma_{b_1}^2, \tag{10}$$

thus dividing our minimand into two parts, due respectively to the bias arising from the neglect of terms of third and higher orders, and to the usual sampling errors.

By the usual least-square theory, the sampling variance of b_1 is

$$\sigma_{b_1}^2 = \mu\sigma^2, \tag{11}$$

where μ is the cofactor of the central element in a, divided by a, that is,

$$\mu = (a_0a_4 - a_2^2)/a. \tag{12}$$

Since μ is of the order of g^{-2}, we may reduce the sampling variance as much as we please by taking the values of x sufficiently far removed from ξ. If $f(x)$ is definitely known to be only of the second degree, a wide dispersion of the desirable values of x is thus indicated, since in this case $E\delta b_1 = 0$, as appears by taking the expectation of each term in (7). But if, as will usually be the case, $f(x)$ has terms of higher orders than the second, an excessively wide dispersion may increase the bias $E\delta b_1$ to such an extent as to render the quadratic approximation inapplicable.

In taking the expectation of each term of (7) and then solving for $E\delta b_1$ we obtain, since $E\gamma_k = 0$ according to the definition of γ_k, and because $E\Delta = 0$, a result of the form

$$E\delta b_1 = B_3\beta_3 + B_4\beta_4 + B_5\beta_5 + \cdots. \tag{13}$$

We shall call B_3, B_4, and B_5 respectively the cubic, quartic and quintic components of the bias, or simply biases. If we denote by λ, μ, ν, the ratios to a of the cofactors of the second column of a, so that

$$\lambda a_1 + \mu a_2 + \nu a_3 = 1, \tag{14}$$

we shall have for the components of bias,

$$\begin{aligned} B_3 &= \lambda a_3 + \mu a_4 + \nu a_5 \\ B_4 &= \lambda a_4 + \mu a_5 + \nu a_6 \\ B_5 &= \lambda a_5 + \mu a_6 + \nu a_7, \end{aligned} \tag{15}$$

and so forth. Since λ, μ, and ν are of respective orders -1, -2 and -3 in a multiplier g of all the values of x, B_3 is of order 2, B_4 is of order 3, and the higher biases are of higher orders. Thus if we begin with any particular distribution of x and apply a sufficiently small multiplier g, we can make the quartic bias negligible in comparison with the cubic, the quintic in comparison with the quartic, and so forth, provided none of these biases is zero. But in reducing g we increase the sampling variance, which is of the order of g^{-2}.

Under these conditions it is reasonable to consider what types of distribution having a fixed value of the sampling variance make the cubic bias a minimum in absolute value; then if there is more than one distribution of this kind, to seek among them a class minimizing the absolute value of the total of cubic and

quartic biases; and among these a class minimizing the absolute value of the total of cubic, quartic and quintic biases, with the modified meaning of the quintic bias taking account of δb_2.

3. The cubic and quartic biases. We find, somewhat unexpectedly, that there exists a class of distributions of x for which the cubic bias is actually zero. To exemplify this we need give the variable no more than three different values, which we may call x, y and z, and we may assign to them the arbitrary frequencies k, m, n of experiments ($k + m + n = N$). If we put

$$P = \begin{vmatrix} 1 & 1 & 1 \\ x & y & z \\ x^2 & y^2 & z^2 \end{vmatrix} = (x - y)(y - z)(z - x), \tag{16}$$

and consider a matrix of three rows and N columns, of which k columns are identical with the first column of P, m with the second, and n with the third, it is evident that the sum of the squares of the three-rowed determinants in this matrix is $kmnP^2$. But this sum of squares is also equal to the determinant formed from the sums of products of the three rows, and this is a (formula (8)). Thus $a = kmnP^2 \neq 0$, since x, y, z are all different. Together with the foregoing $3 \times N$ matrix consider another,

$$\left\| \begin{matrix} 1 \cdots & 1 \cdots\cdots & 1 \cdots \\ x^2 \cdots & y^2 \cdots\cdots & z^2 \cdots \\ x^3 \cdots & y^3 \cdots\cdots & z^3 \cdots \end{matrix} \right\|, \tag{17}$$

having k columns identical with that first written, m identical with the second written, and n identical with the third. The only non-vanishing three-rowed determinants in this matrix are formed of these three different columns, and equal $(xy + yz + zx)P$; there are kmn of them. The sum of products of corresponding three-rowed determinants in the two matrices is therefore $kmnP^2(xy + yz + zx)$. But this sum is also equal to the determinant, formed from the sums of products of corresponding rows,

$$\begin{vmatrix} a_0 & a_2 & a_3 \\ a_1 & a_3 & a_4 \\ a_2 & a_4 & a_5 \end{vmatrix},$$

which, by (15), equals $-aB_3$. It follows that

$$-B_3 = xy + yz + zx. \tag{18}$$

There are many real solutions of the equation

$$xy + yz + zx = 0, \tag{19}$$

with the three values all different, for example $-2, 3, 6$. If we assign such values to our variable, and an arbitrary number of experimental determinations to each of these values, the cubic bias B_3 will be zero.

It will be noticed that such a solution cannot have zero for one of the values. If, for example, $z = 0$ in (19), then x or y must also vanish, in violation of the condition that there must be at least three distinct values. Moreover a solution cannot be symmetrical about zero; if $x + y = 0$ it follows from (19) that $x = y = 0$ A solution may or may not be symmetrical about a value other than zero. The values $3 - 2\sqrt{3}$, $(3 - \sqrt{3})/2$, $\sqrt{3}$ satisfy the equation and are in arithmetic progression, while the solution $-2, 3, 6$ is asymmetrical.

If we modify (17) by replacing the cubes of the variables by their fourth powers, and apply the same procedure to the modified matrix, we find that

$$B_4 = -(x + y)(y + z)(z + x). \tag{20}$$

Thus there exist sets of three distinct real values making the quartic bias vanish, for example any set for which $x + y = 0$; but no such set can at the same time nullify the cubic bias (18). Since it is ordinarily more important for the cubic than for the quartic bias to vanish, distributions nullifying (20) are not in general to be recommended. But in exceptional cases it may be known that β_3 is zero, or very small in comparison with β_4, and then the vanishing of B_4 is a more valuable property than that of B_3. It will be shown that no distribution of three or more values exists such that both the cubic and quartic components of bias are zero.

Let us denote by D_p the p-rowed determinant having a_{i+j-2} as the element in its ith row and jth column. Thus D_3 is the same determinant which we have in (8) called a, and

$$D_4 = \begin{vmatrix} a_0 & a_1 & a_2 & a_3 \\ a_1 & a_2 & a_3 & a_4 \\ a_2 & a_3 & a_4 & a_5 \\ a_3 & a_4 & a_5 & a_6 \end{vmatrix}. \tag{21}$$

For every distribution, every $D_p \geq 0$; and a necessary and sufficient condition that a distribution have p or more distinct values is that D_p be *greater* than zero. [7, p. 362]. If D_p is positive, so is each of its principal minors. In particular, since we are requiring at least three values in a distribution, $D_3 = a > 0$, and therefore

$$a_2a_4 - a_3^2 > 0, \tag{22}$$

and

$$a_0a_4 - a_2^2 > 0. \tag{23}$$

We shall now consider distributions for which the cubic bias B_3 is zero, and consequently, by (15),

$$\lambda a_3 + \mu a_4 + \nu a_5 = 0, \tag{24}$$

and expand D_4. From the definition of λ, μ, ν, we have

$$\lambda a_2 + \mu a_3 + \nu a_4 = 0. \tag{25}$$

Multiply the last row of the determinant (21) by ν, and add to it λ times the second row and μ times the third. The last row is thus, by (14), (25), (24) and (15) transformed into

$$1 \quad 0 \quad 0 \quad B_4,$$

while the determinant has been multiplied by ν. Let this new determinant be expanded with respect to its last row. The cofactor of the first element 1 is

$$G = -\begin{vmatrix} a_1 & a_2 & a_3 \\ a_2 & a_3 & a_4 \\ a_3 & a_4 & a_5 \end{vmatrix}.$$

Let the last row of this determinant be multiplied by ν, an operation having the effect of multiplying the whole determinant by ν; and let λ times the first row and μ times the second row then be added to the last. The last row is thus, by (14), (25) and (24) reduced to

$$1 \quad 0 \quad 0.$$

Hence

$$\nu G = -(a_2 a_4 - a_3^2),$$

and consequently

$$\begin{aligned} \nu^2 D_4 &= \nu(aB_4 + G) \\ &= \nu a B_4 - (a_2 a_4 - a_3^2). \end{aligned} \tag{26}$$

Since the first member of this equation is positive or zero, (22) shows that it is impossible that B_4 should equal zero when $B_3 = 0$ as we have assumed. That is,

Either the cubic or the quartic bias of every distribution having three or more distinct values must be different from zero.

If ν were zero, (26) would contradict (22). Hence $\nu \neq 0$. With every distribution of x there is associated another obtained from it by changing the sign of each value of x. Such a pair of distributions we shall call *opposite*. When we pass from a distribution to its opposite, the power-sums a_k remain unchanged when k is even and change only in sign when k is odd. Since a is always positive, and since

$$\nu = (a_1 a_2 - a_0 a_3)/a, \tag{27}$$

ν has opposite signs and the same absolute value for opposite distributions. The conclusions to be reached shortly will be equally valid for a distribution and its opposite, and in reaching them we may assume $\nu > 0$. It will then follow from (22) and (26) that $B_4 > 0$.

4. Distributions nullifying cubic bias with minimum quartic bias. We can now prove the following theorem:

Among distributions for which the cubic bias vanishes and the standard error of b_1 has a fixed value, those for which the quartic bias is a minimum have exactly three distinct values of the variable. These values satisfy the equation

$$xy + yz + zx = 0. \tag{28}$$

Since the standard error σ of a single observation is not affected by the distribution chosen for x, fixation of the standard error of b_1 is equivalent by (11) to fixation of the value of the expression μ given by (12). We suppose therefore that μ has some fixed positive value and that $B_3 = 0$. Since μ, B_3 and B_4 also not involve the distribution of x excepting through the power-sums $a_0, a_1, \cdots, a_6$, we may treat these power-sums as the independent variables in trying to make B_4 a minimum. Their region of variation is limited by the inequalities referred to in the preceding section,

$$D_1 = a_0 > 0, \qquad D_2 > 0, \qquad D_3 = a > 0, \qquad D_4 \geq 0.$$

The inequalities $D_p \geq 0$ for $p > 4$ involve power-sums of orders higher than the sixth and are irrelevant to our purpose.

The definition (8) of a shows that it is independent of a_5 and a_6; consequently λ, μ, and ν are also. According to (15), B_3 involves a_5 but not a_6; while of all the expressions we have considered, only B_4 and D_4 are functions of a_6. Therefore when $a_0, a_1, \cdots, a_5$ are given any definite values, a_6 may be chosen to make B_4 a minimum without any regard to the fixed values of μ and B_3. Now (15) shows that B_4 is a linear function of a_6 with a coefficient which, at the end of the last section, we have proved not to be zero and assumed positive. Thus B_4, which is also positive, is an increasing function of a_6. Its minimum will correspond to the least value of a_6 consistent with the condition $D_4 \geq 0$. But (21) shows that D_4 is also a positive linear function of a_6 with a positive coefficient, a. The minimum of a_6, and therefore that of B_4, require therefore that $D_4 = 0$. But $D_4 = 0$ is exactly the condition that there should be no more than three distinct values in the distribution. Since there must be at least three distinct values, and since if there are only three they must satisfy (19), the theorem is proved.

The minimum value of B_4 with respect to variations of a_6 when $B_3 = 0$ may be found by putting $D_4 = 0$ in (26). Designating this minimum by b and using (27) we have

$$b = \frac{a_2 a_4 - a_3^2}{a_1 a_2 - a_0 a_3}, \tag{29}$$

where the numerator is intrinsically positive, and the denominator is positive for the class of distributions we are now considering, though we might equally well consider the opposite distributions, for which it is negative. We have also from (20),

$$(x + y)(y + z)(z + x) = -b. \tag{30}$$

Substituting for each of these binomials its value as given by (28), we may write this in the simpler form

$$xyz = b > 0. \tag{31}$$

It was shown at the beginning of section 3 that when there are only three values in the distribution, with frequencies k for x, m for y, and n for z,

$$a = kmnI^2 = kmn(x - y)^2(y - z)^2(z - x)^2. \tag{32}$$

The first two rows of (17) form a matrix such that the sum of the squares of its two-rowed determinants is

$$mn(y^2 - z^2)^2 + nk(z^2 - x^2)^2 + km(x^2 - y^2)^2. \tag{33}$$

Since this is equal to the determinant of the sums of products of the rows, namely

$$\begin{vmatrix} a_0 & a_2 \\ a_2 & a_4 \end{vmatrix},$$

it follows from (12), (32) and (33) that

$$\mu = \frac{(y + z)^2}{k(x - y)^2(x - z)^2} + \frac{(x + z)^2}{m(x - y)^2(y - z)^2} + \frac{(x + y)^2}{n(x - z)^2(y - z)^2}. \tag{34}$$

It is desired to minimize this expression, which is the factor of the variance that is independent of the accuracy of the individual observations, while holding $b = xyz$ fixed; or to minimize b while holding μ fixed. In either case the values of x, y and z are to be chosen to satisfy (28). The relations established by the solution of either of these virtually equivalent problems will fix x, y, and z except for a factor of proportionality, which must then be adjusted to provide a balance as satisfactory as possible between random errors and bias.

5. The quintic bias. Effect of δb_2. With any distribution determined in this way will be associated its opposite distribution, which will have the same minimizing properties so far as the variance and the cubic and quartic components of bias are concerned. The appropriate choice between these two opposite distributions will in general involve the quintic component of the bias. At this point we must, for the first time, take account of the errors in the denominator b_2 of x_0.

Since b_1 converges stochastically to Eb_1, and b_2 to Eb_2, the error $x_0 = -\frac{1}{2}b_1/b_2$ converges stochastically (for large samples) to $-\frac{1}{2}Eb_1/Eb_2$. By keeping our

values of x close enough to ξ we may insure that Eb_2 differs as little as we please from β_2, and hence that the series

$$\frac{Eb_1}{Eb_2} = \frac{Eb_1}{\beta_2 + E\delta b_2} = \frac{Eb_1}{\beta_2}\left\{1 - \frac{E\delta b_2}{\beta_2} + \frac{(E\delta b_2)^2}{\beta_2^2} - \cdots\right\}$$

converges rapidly. Let us rearrange this series after inserting for Eb_1 and $E\delta b_2$ their values, so as to obtain a series in ascending powers of a common multiplier g which may be applied to the values of x. We recall that in the expression (13) for Eb_1, B_3 is of the second order in g, B_4 is of the third order, B_5 is of the fourth order, and so forth. In the same way, we find that

$$E\delta b_2 = C_3\beta_3 + C_4\beta_4 + \cdots,$$

where

$$C_3 = \frac{1}{a}\begin{vmatrix} a_0 & a_1 & a_3 \\ a_1 & a_2 & a_4 \\ a_2 & a_3 & a_5 \end{vmatrix}$$

is of the first order, C_4 is of the second order, and so forth. Thus in

$$\beta_2\frac{Eb_1}{Eb_2} = B_3\beta_3 + (B_4\beta_4 - B_3C_3\beta_3^2/\beta_2) + (B_5\beta_5 - B_4C_3\beta_3\beta_4/\beta_2 - B_3C_4\beta_3\beta_4/\beta_2 + B_3C_3^2\beta_3^4/\beta_2^2) + \cdots,$$

the first term is of the second order, those in the first parentheses are of the third order, those in the second parentheses are of fourth order, and the remaining terms are of higher orders.

We have seen that we can choose distributions for which $B_3 = 0$. In this way we get rid of the second-order term and reduce the third-order terms to $B_4\beta_4$. We shall in the next two sections show how, under various conditions, to select from among the distributions for which $B_3 = 0$ an opposite pair for each of which $|B_4|$ is a minimum. In choosing between these two opposite distributions, the criterion we shall adopt is that the terms of third order and those of fourth order shall have opposite signs; for while the fourth-order terms may be made much smaller than those of third order in absolute value, still it is desirable that they should offset them, in order to reduce the error. The terms of third and of fourth orders reduce respectively for $B_3 = 0$ to $B_4\beta_4$ and to $B_5\beta_5 - {}^4B_4C_3\beta_3\beta_4/\beta_2$. Our criterion is that these are to have opposite signs, and consequently that

$$B_4\beta_2\beta_4(B_5\beta_2\beta_5 - B_4C_3\beta_3\beta_4) < 0.$$

We shall however modify this criterion whenever σ is not negligibly small. A more precise criterion will be obtained by expanding x_0^2 in a series of powers of δb_2, taking the expectation term by term, and reducing the moments thus obtained of orders higher than the second to those of first and second orders by

means of the theory of the bivariate normal distribution of b_1 and b_2. It is then necessary to make some assumption regarding the order of magnitude of x, y and z relatively to N in order to assemble terms of like magnitude in a criterion resembling that above but involving σ. The appropriate balance indicated by the results of the next two sections calls for x, y and z to be of the order of $N^{-1/8}$. This leads to the following criterion:

$$\beta_2(B_4B_5\beta_2\beta_4\beta_5 - B_4^2C_3\beta_3\beta_4^2 - C_3\beta_3\mu\sigma^2) < 0.$$

We have seen that $B_4 = b = xyz$. To evaluate C_3 and B_5, which latter may in accordance with (15) be written

$$B_5 = -\frac{1}{a}\begin{vmatrix} a_0 & a_2 & a_5 \\ a_1 & a_3 & a_6 \\ a_2 & a_4 & a_7 \end{vmatrix},$$

we proceed as in section 3, replacing the second row of (17) by the first powers to obtain C_3, and replacing the third row of (17) by the fifth powers of x, y and z to obtain B_5. In this way we find

$$C_3 = \frac{1}{P}\begin{vmatrix} 1 & 1 & 1 \\ x & y & z \\ x^3 & y^3 & z^3 \end{vmatrix}, \qquad B_5 = -\frac{1}{P}\begin{vmatrix} 1 & 1 & 1 \\ x^2 & y^2 & z^2 \\ x^5 & y^5 & z^5 \end{vmatrix}.$$

Letting Σx, Σx^2yz, etc. stand for the symmetric functions of x, y and z of which one term is written in each case after Σ, we may reduce these expressions to

$$C_3 = \Sigma x,$$
$$B_5 = -\Sigma x^3y - \Sigma x^2y^2 - 2\Sigma x^2yz.$$

With the help of (28) and (31) we find

$$\Sigma x^2yz = xyz\Sigma x = b\Sigma x,$$
$$\Sigma x^2y^2 = (\Sigma xy)^2 - 2\Sigma x^2yz = -2b\Sigma x,$$
$$\Sigma x^3y = \Sigma xy\Sigma x^2 - \Sigma x^2yz = -b\Sigma x.$$

Therefore $B_5 = b\Sigma x$. Substituting these values for B_4, C_3 and B_5 in the last inequality gives the rule:

Choose that one of a pair of opposite distributions for which

$$(35) \qquad (x + y + z)\beta_2\{b^2\beta_4(\beta_2\beta_5 - \beta_3\beta_4) - \beta_3\mu\sigma^2\} < 0.$$

It will be remembered that β_2 is negative for a maximum of $f(x)$, positive for a minimum. The other β's can only be estimated from preliminary experimentation, or possibly in particular cases from general knowledge or theory.

Quite different algebraic methods are appropriate to minimizing μ with a fixed

b according to the limitations to be placed on the frequencies k, m, n; the methods leading very simply to a solution in one case involve trouble some complications in another. We shall deal with two of the leading cases.

6. The case of equal frequencies. Some experimental situations call for equal frequencies for all values of the variable. If $k = m = n$, then $a_0 = N = 3n$. Let $a_j' = a_j/n$. Then $a_0' = 3$ and $a_1' = \Sigma x$. Inasmuch as

$$\Sigma xy = 0 \qquad \text{and} \qquad xyz = b, \tag{36}$$

we may express a_2', a_3' and a_4' as functions of a_1' and b as follows:

$$a_2' = \Sigma x^2 = (\Sigma x)^2 - 2\Sigma xy = a_1'^2.$$

$$a_3' = \Sigma x^3 = (\Sigma x)^3 - 3\Sigma x^2 y - 6xyz;$$

and since $\Sigma x^2 y = \Sigma x \Sigma xy - 3xyz$ we have from (36),

$$a_3' = a_1'^3 + 3b.$$

We have also

$$a_4' = \Sigma x^4 = (\Sigma x)^4 - 4\Sigma x^3 y - 6\Sigma x^2 y^2 - 12\Sigma x^2 yz,$$

and since

$$\Sigma x^3 y = \Sigma xy \Sigma x^2 - \Sigma x^2 yz, \qquad \Sigma x^2 yz = xyz\Sigma x = a_1' b,$$

$$\Sigma x^2 y^2 = (\Sigma xy)^2 - 2\Sigma x^2 yz = -2a_1' b,$$

it follows that

$$a_4' = a_1'^4 + 4a_1' b.$$

Therefore

$$a = n^3 \begin{vmatrix} 3 & a_1' & a_1'^2 \\ a_1' & a_1'^2 & a_1'^3 + 3b \\ a_1'^2 & a_1'^3 + 3b & a_1'^4 + 4a_1' b \end{vmatrix}$$

Upon subtracting a_1' times the second column from the third, and a_1' times the first from the second, this becomes

$$a = n^3 b \begin{vmatrix} 3 & -2a_1' & 0 \\ a_1' & 0 & 3 \\ a_1'^2 & 3b & a_1' \end{vmatrix} = -n^3 b(4a_1'^3 + 27b).$$

Also,

$$a_0 a_4 - a_2^2 = n^2\{3(a_1'^4 + 4a_1' b) - (a_1'^2)^2\} = 2n^2(a_1'^4 + 6a_1' b).$$

Hence, by (12),

$$\mu = \frac{a_0 a_4 - a_2^2}{a} = -\frac{2}{nb}\,\frac{a_1'^4 + 6a_1' b}{4a_1'^3 + 27b}. \tag{37}$$

Differentiating with respect to a_1' to find a minimum, we obtain

$$0 = (4a_1'^3 + 27b)(4a_1'^3 + 6b) - 12a_1'^2(a_1'^4 + 6a_1'b) = 4a_1'^6 + 60a_1'^3 + 162b^2.$$

The minimum of μ, for b fixed, and satisfying the condition $4a_1'^3 + 27b < 0$, which is equivalent to $a > 0$ since we assume $b > 0$, is attained when $a_1'^3 = bq$ where q is the numerically greater root of the equation $4q^2 + 60q + 162 = 0$; that is,

$$q = -(15 + \sqrt{63}/2) = -11.468\ 626\ 97.$$

The elementary symmetric functions of the values x, y, z composing the distribution are

$$\Sigma x = a_1' = (bq)^{1/3}, \qquad \Sigma xy = 0, \qquad xyz = b.$$

Hence x, y and z must be the roots of the equation in u,

$$u^3 - (bq)^{1/3}u^2 - b = 0. \tag{38}$$

If we put $u = (bq)^{1/3}v$,

$$v^3 - v^2 + q^{-1} = 0.$$

Calculation gives approximately

$$q^{-1} = -.087\ 194\ 396, \text{ and for the roots of the equation in } v, \tag{39}$$
$$.2628, \qquad -.3729, \qquad -.8899,$$

numbers which are therefore proportional to the values of the variable that should be chosen when the frequencies must be equal. If any values x, y, z proportional to these are used, the value (37) of μ is

$$\mu' = -\frac{6}{N}\frac{q+6}{4q+27}q^{1/3}b^{-2/3} \tag{40}$$

and is the minimum consistent with any fixed value b of xyz.

Choice of the factor of proportionality will involve a compromise between the criteria of minimum sampling variance and minimum bias. If we ignore components of bias of orders higher than the fourth and recall (10) and (11) it will appear that the appropriate combined criterion is that

$$b^2\beta_4^2 + \mu\sigma^2 \tag{41}$$

shall be a minimum. Putting for μ its value μ' from (40) and differentiating with respect to b gives

$$2\beta_4^2 b + \frac{4\sigma^2 q^{1/3}}{N}\frac{q+6}{4q+27}b^{-5/3} = 0,$$

or

$$b = b' = \left(-\frac{2\sigma^2}{N\beta_4^2}\frac{q+6}{4q+27}q^{1/3}\right)^{3/8}.$$

The product of the three roots (39) is $-q^{-2}$. Numbers proportional to these and having the product b' will be obtained by multiplying them by $-b'q^{-2}$, that is, by

$$\left(\frac{\sigma^2}{N\beta_1^2}\right)^{1/4}\left(2\,\frac{q+6}{4q+27}\right)^{1/4}(-q)^{3/2} = 2.3318\left(\frac{\sigma^2}{N\beta_1^2}\right)^{1/4}.$$

Multiplying (39) by 2.3318 gives numbers

$$.6128, \qquad -.8695, \qquad -2.0751, \tag{42}$$

which must still be multiplied by $\pm\,[\sigma^4/(N\beta_1^2)]^{1/4}$ to give the set minimizing $E\delta b_1^2$. The ambiguous sign is to be fixed according to the rule at the end of the last section. Thus we arrive finally at the conclusion:

If the numbers of observations are required to be the same for all the values of the variable used, these values should for greatest efficiency deviate from the estimated maximizing value by the products of the three numbers (42) by

$$\pm\left(\frac{\sigma^2}{N\beta_1^2}\right)^{1/4}, \tag{43}$$

choosing the ambiguous sign so as to satisfy (35).

The product b' of the three values is to be substituted for b in (40) and (35), and the value of μ thus obtained from (40) is also to be substituted in (35). These substitutions yield

$$(x+y+z)\beta_2\beta_4(\beta_2\beta_5 - 4\beta_3\beta_4) < 0$$

as the criterion for choosing the sign in (43).

The expectation of the square of the error in the estimate of the value x_0 of ξ is, according to (9) and (10), given approximately by the ratio of (41) to β_2^2, and it is this that will be a minimum when the foregoing rule is followed. The minimum of (41) is obtained by replacing b by b' in (40) and (41), and substituting (40) for μ in (41). This gives

$$E(\delta b_1)^2 = 4\left(\frac{2}{N}\,\frac{q+6}{4q+27}\right)^{3/4}(-q)^{1/4}\beta_4^{1/2}\sigma^{3/2};$$

that is,

$$E(\delta b_1)^2 = 4.889\,N^{-3/4}\beta_4^{1/2}\sigma^{3/2}. \tag{44}$$

7. Adjustable frequencies. If the total number N of observations to be made can be distributed freely among the values of the variable, the efficiency of the experiment can be increased by a proper selection of the individual frequencies

k, m, n along with the corresponding values x, y, z. We shall choose these six unknowns, subject to the three conditions[4]

$$k + m + n = N, \tag{45}$$

$$xy + xz + yz = 0, \tag{46}$$

$$xyz = -b, \tag{47}$$

to minimize μ. The last condition fixes the quartic bias, the preceding one expresses the vanishing of the cubic bias. It is of course understood that k, m, n are all positive, and we shall, as before, suppose initially that b is positive. No two of x, y, z can be equal, and it follows that none of them, or of the sums of two of them, can be zero while satisfying the second condition. We shall lose no generality in assuming that

$$x > y > 0 > z. \tag{48}$$

Furthermore, it is easy to see that $x + y$, $x + z$, and $y + z$ are all positive Therefore the quantities

$$r = \frac{y+z}{(x-y)(x-z)}, \quad s = \frac{x+z}{(x-y)(y-z)}, \quad t = \frac{x+y}{(x-z)(y-z)} \tag{49}$$

are all positive. From (34) we have

$$\mu = \frac{r^2}{k} + \frac{s^2}{m} + \frac{t^2}{n}. \tag{50}$$

The values of k, m, n making this a minimum while themselves subject to the limitation that their sum is N must if they were continuous positive variables be proportional to r, s and t. Of course the frequencies are integers, but we are supposing N large, so that the values found by differentiation will be close approximations, and we shall disregard this complication. Put therefore

$$r = k\rho, \qquad s = m\rho, \qquad t = n\rho, \tag{51}$$

where ρ is a multiplier which evidently is not zero. If we use these equations to eliminate r, s, t from μ we obtain, with the help of (45), $\mu = N\rho^2$. But if we use them to eliminate k, m, n from (50) we have instead,

$$\mu = (r + s + t)\rho.$$

Now from (49),

$$r + t = s, \tag{52}$$

[4] The condition (47) is here used instead of (31), from which it differs by the introduction of the negative sign, because it simplifies the argument of this section slightly to have the quantities (49) positive. There is no essential difference, since we are seeking a pair of opposite distributions.

so that $\mu = 2s\rho$. Therefore $N\rho = 2s$, and finally $\mu = 4s^2/N$. Therefore μ is a minimum when the positive quantity s is a minimum. In the expression (49) for s we substitute from (46) and (47)

$$x + z = -xz/y = b/y^2, \tag{53}$$
$$(x - y)(y - z) = (x + z)y - xz - y^2 = 2b/y - y^2,$$

so that

$$s = \frac{b}{y(2b - y^3)}. \tag{54}$$

Since y, s and b are positive, this shows that $y^3 < 2b$. The value of y on the interval from 0 to $2b$ making s a minimum is found by differentiation to be $2^{-1/3}b^{1/3}$. Substituting this in (53) and (47) gives

$$x + z = 2^{2/3}b^{1/3}, \qquad xz = -2^{1/3}b^{2/3},$$

whence

$$x = (b/2)^{1/3}(1 + \sqrt{3}), \qquad y = (b/2)^{1/3}, \qquad z = (b/2)^{1/3}(1 - \sqrt{3}). \tag{55}$$

From (45), (51) and (52) it is seen that $k + n = m = N/2$. Thus half the total observations are to be concentrated on the middle value. From (51) and (49) we have also

$$\frac{k}{n} = \frac{r}{t} = \frac{y^2 - z^2}{x^2 - y^2},$$

wherefore

$$k = \frac{N}{2}\frac{y^2 - z^2}{x^2 - z^2}, \qquad n = \frac{N}{2}\frac{x^2 - y^2}{x^2 - z^2}.$$

With (55) this shows that

$$\begin{aligned} k &= N(2 - \sqrt{3})/8 \quad m = N/2, \qquad n = N(2 + \sqrt{3})8 \\ &= .03349\ N, \qquad\qquad\qquad = .46651\ N. \end{aligned} \tag{56}$$

We have seen that $\mu = 4s^2/N$. Substituting in (54) the value found for y gives $s = 2^{4/3}b^{-1/3}/3$. Therefore the minimum of μ for a fixed value of b is

$$\mu = (16/9N)(2/b)^{2/3} \tag{57}$$

Inserting this in the expression (41) for the total expectation of the squared error and then differentiating with respect to b gives

$$b = 2^{7/4}3^{-9/8}N^{-3/8}\beta_4^{-3/4}\sigma^{3/4}. \tag{58}$$

When this value is given to b, (41) becomes

$$3.8207N^{-3/4}\beta_4^{1/2}\sigma^{3/2}. \tag{59}$$

The greater efficiency of experiments with the frequencies (56) and the correspondingly adjusted values x, y, z, in comparison with the case in which the frequencies must be equal, corresponds to the smaller coefficient in (59) than in (44). To obtain as great accuracy with equal frequencies as with adjusted ones it is necessary to have more observations, in a ratio obtained by equating (59) with (44) after inserting different symbols for N in the two cases. In this way it is found that the number of observations required with efficient distribution of the frequencies is almost exactly 72 per cent of the number required when the frequencies are equal, if the values x, y, z are in each case given their most efficient values.

Substituting (58) in (55) gives the numbers

$$2.1520, \quad .7877, \quad -.2110, \tag{60}$$

multiplied by (43), with a change of signs if necessary to satisfy (35), as the values x, y, z of the variable to be used. The more concentrated character of this distribution with adjustable frequencies is emphasized by the small proportion, less than $3\frac{1}{2}$ per cent, of the frequencies (56) that pertains to the value most remote from the tentative maximizing value.

When (58) is substituted in (57) and, with the result, in (35), this inequality reduces to exactly the same form as that obtained in the preceding section for fixing the sign of (43).

8. Introduction to the two-variable problem. Functions of two or more variables are of greater practical importance than functions of one variable. The recent work on factorial experiments [8] makes it clear that in the experimental determination of maxima of functions of several variables, considerable improvements are possible over the practice of trying the effect of variations in only one variable at a time while holding the others constant. It seems likely that the methods worked out in the previous sections for experimenting with one variable are capable of generalization. However certain difficulties enter which have not yet been surmounted. The object of the present section is to indicate something of the nature of the problem of extending the foregoing results to two variables, x and y.

Let us suppose that a quadratic regression equation,

$$Z = b_{00} + b_{10}x + b_{01}y + \tfrac{1}{2}(b_{20}x^2 + 2b_{11}xy + b_{02}y^2),$$

will be fitted by least squares to observations of $z = f(x, y)$ based on N combinations of x and y, each of which represents a point in a plane. Since there are six coefficients to be determined, there must be at least six distinct points

$(x_1, y_1), \cdots, (x_6, y_6)$. The coefficients in the normal equations may be written $a_{jk} = Sx^j y^k$, so that $a_{00} = N$. The determinant

$$a = \begin{vmatrix} a_{00} & a_{10} & a_{01} & a_{20} & a_{11} & a_{02} \\ a_{10} & a_{20} & a_{11} & a_{30} & a_{21} & a_{12} \\ a_{01} & a_{11} & a_{02} & a_{21} & a_{12} & a_{03} \\ a_{20} & a_{30} & a_{21} & a_{40} & a_{31} & a_{22} \\ a_{11} & a_{21} & a_{12} & a_{31} & a_{22} & a_{13} \\ a_{02} & a_{12} & a_{03} & a_{22} & a_{13} & a_{04} \end{vmatrix}$$

must not vanish. Let the function under investigation be

$$f(x, y) = \Sigma\Sigma\beta_{jk} x^j y^k/(j + k)!,$$

and suppose that $\beta_{10} = 0 = \beta_{01}$, so that the origin is the point sought at which the first derivatives vanish. We shall assume that

$$\beta = \beta_{20}\beta_{02} - \beta_{11}^2 > 0, \qquad \beta_{20} < 0,$$

implying a definite maximum. The estimates x_0, y_0 of the maximizing (or minimizing) values obtained by differentiating Z are

$$x_0 = (b_{11}b_{01} - b_{02}b_{10})/b, \qquad y_0 = (b_{11}b_{10} - b_{20}b_{01})/b,$$

where

$$b = b_{20}b_{02} - b_{11}^2.$$

For large samples and values of x and y taken not too far from the origin, b will approximate to β, and x_0 and y_0 respectively to

$$(\beta_{11}b_{01} - \beta_{02}b_{10})/\beta, \qquad (\beta_{11}b_{10} - \beta_{20}b_{01})/\beta.$$

Some means is needed of combining into one the two desiderata of minimizing the errors x_0 and y_0. A combined measure of these deviations is

$$\beta_{20}x_0^2 + 2\beta_{11}x_0y_0 + \beta_{02}y_0^2.$$

This expression is constant except for terms of higher order when x_0 and y_0, while remaining small, vary in such a way that $f(x, y)$ maintains a constant value. Substituting in it the approximate values of x_0 and y_0 gives β^{-1} times

$$\beta_{02}b_{10}^2 - 2\beta_{11}b_{10}b_{01} + \beta_{20}b_{01}^2.$$

The expectation of this measure of error may be separated into two parts by means of the formulae for the variances and covariance,

$$\sigma_{b_{10}}^2 = Eb_{10}^2 - (Eb_{10})^2, \qquad \sigma_{b_{10}b_{01}} = Eb_{10}b_{01} - (Eb_{10})(Eb_{01}), \quad \text{etc.}$$

One of these parts is a generalized sampling variance,

$$\beta_{02}\sigma_{b_{10}}^2 - 2\beta_{11}\sigma_{b_{10}b_{01}} + \beta_{20}\sigma_{b_{01}}^2,$$

and tends to zero with order N^{-1} as N increases provided the values (x_k, y_k) are fixed. The other part,

$$\beta_{02}(Eb_{10})^2 - 2\beta_{11}(Eb_{10})(Eb_{01}) + \beta_{20}(Eb_{01})^2, \tag{61}$$

is a bias which does not tend to zero as N increases, but which may be kept arbitrarily small, at the expense of the sampling variance, by restricting the values (x_k, y_k) to be sufficiently small. This expression is a negative definite quadratic form in Eb_{10} and Eb_{01}, and therefore cannot be zero unless both these components of bias vanish separately.

We may proceed as in paragraph 2 to express Eb_{10} and Eb_{01} in terms of the coefficients of $f(x, y)$ of orders higher than the second, among which those of third order will be of leading importance. In this way it may be shown that, if we neglect terms in $f(x, y)$ of orders higher than the third, Eb_{10} and Eb_{01} are given by the ratios to a constant multiple of a of determinants obtained from a by replacing respectively the second and the third columns by the column

$$\begin{aligned}
&\beta_{30}a_{30} + 3\beta_{21}a_{21} + 3\beta_{12}a_{12} + \beta_{03}a_{03}\\
&\beta_{30}a_{40} + 3\beta_{21}a_{31} + 3\beta_{12}a_{22} + \beta_{03}a_{13}\\
&\beta_{30}a_{31} + 3\beta_{21}a_{22} + 3\beta_{12}a_{13} + \beta_{03}a_{04}\\
&\beta_{30}a_{50} + 3\beta_{21}a_{41} + 3\beta_{12}a_{32} + \beta_{03}a_{23}\\
&\beta_{30}a_{41} + 3\beta_{21}a_{32} + 3\beta_{12}a_{23} + \beta_{03}a_{14}\\
&\beta_{30}a_{32} + 3\beta_{21}a_{23} + 3\beta_{12}a_{14} + \beta_{03}a_{05}.
\end{aligned}$$

It is desirable to select a distribution of points (x_k, y_k) such that these components of bias will vanish, no matter what may be the values of β_{30}, β_{21}, β_{12} and β_{03}. For this it is necessary and sufficient that all the determinants vanish that are obtained from these two by replacing the column written above by the terms in it that multiply any one of the four β_{jk}'s. The single-variable analogy suggests using a distribution having the smallest possible number of points, which in this case is six. Let us now take $N = 6$. The eight determinants will all be multiples of

$$P = \begin{vmatrix} 1 & x_1 & y_1 & x_1^2 & x_1y_1 & y_1^2 \\ 1 & x_2 & y_2 & x_2^2 & x_2y_2 & y_2^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_6 & y_6 & x_6^2 & x_6y_6 & y_6^2 \end{vmatrix}.$$

To save space we shall indicate determinants of this character merely by writing a single row without subscripts, thus:

$$P = |\, 1 \quad x \quad y \quad x^2 \quad xy \quad y^2 \,|.$$

If we define

$$A'_{jk} = \begin{vmatrix} 1 & x^j y^k & y & x^2 & xy & y^2 \end{vmatrix},$$
$$A''_{jk} = \begin{vmatrix} 1 & x & x^j y^k & x^2 & xy & y^2 \end{vmatrix},$$

and multiply each of these determinants for which $j + k = 3$ $(j, k = 0, 1, 2, 3)$ by P, columns by columns, we shall have exactly the determinants whose vanishing is the condition for nullification of the cubic bias. If we multiply P by itself in the same way we have $P^2 = a$. Therefore $P \neq 0$. Therefore the required condition is that the distribution satisfy the eight equations

$$A'_{30} = 0, \quad A'_{21} = 0, \quad A'_{12} = 0, \quad A'_{03} = 0,$$
$$A''_{30} = 0, \quad A''_{21} = 0, \quad A''_{12} = 0, \quad A''_{03} = 0,$$

and the inequality $P \neq 0$.

In seeking distributions nullifying the cubic bias we have twelve unknowns $x_1, \cdots, x_6, y_1, \cdots, y_6$ which must satisfy these eight equations. This suggests that we give arbitrary values to four of them and then solve for the other eight by straightforward elimination. Unfortunately, since the eight equations are each of the tenth degree, reducing to the ninth degree when coordinates of two of the points are given numerical values, a straightforward elimination would seem to lead to an equation of degree $9^8 = 43{,}046{,}721$. The number of algebraic operations in performing the elimination, solving the equation for one of the unknowns, substituting back, and solving for the others, would be a large multiple of this number, and would doubtless be sufficient to occupy a large and efficient computing project for many millenniums. At the end of this period it might be found that the roots corresponding to the original arbitrary values chosen were all complex or made $P = 0$, and were therefore unusable. Thus indirect and less elementary methods are called for, and some qualitative investigations of such distributions, if they exist (which is not certain), are in order.

The set of conditions as a whole is invariant under all non-singular homogeneous linear transformations of x and y, as is easily proved by making linear combinations of the columns of each of the determinants A'_{jk}, A''_{jk} and P, and by making linear combinations of these determinants themselves. These linear transformations leave the origin invariant. They have four degrees of freedom, which is exactly the right number to take care of the excess of unknowns over equations. This points to the possible existence of a finite number of fundamental solutions, from which all solutions may be obtained by linear homogeneous transformations. Geometrical properties of the configuration will be represented by invariants under linear transformations. Thus the condition $P \neq 0$ means that the six points must not all lie on any conic section. From this it follows at once that no four of them can lie on a straight line, since this line, with the line through the other two, would constitute a degenerate conic. As a matter of fact, we can go further and prove that no three of the points

may lie on a straight line. In the proof of this and other properties of the distribution it is convenient to use the arbitrariness provided by a linear transformation to pass the axes (which may be oblique) through any two of the six points, and then to adjust the scales of measurement so that the coordinates of these points become (1, 0) and (0, 1), except that one of them might conceivably be the origin. If three points are collinear, their line can be taken to be the x-axis if it passes through the origin, or the line $y = 1$ if it does not. Even with the help provided by such procedures the proofs are rather long, though straightforward. We shall content ourselves here with stating, without proof, the following properties necessary for sets of six points for which $P \neq 0$ and all components of the cubic bias vanish:

No three of the points can lie on a straight line

No two straight lines through the origin can contain four of the points.

No four of the points can lie on the vertices of a parallelogram.

The set cannot consist of the origin and the vertices of a regular pentagon with center at the origin.

These conditions have been established by calculations of a rather straightforward and laborious sort, too long to be reproduced.

If $z_k = x_k + iy_k$ and $\bar{z}_k = x_k - iy_k$, the conditions $P \neq 0$, $A'_{1k} = 0 = A''_{1k}$, may be written

$$\left| \begin{matrix} 1 & z & \bar{z} & z^2 & z\bar{z} & \bar{z}^2 \end{matrix} \right| \neq 0, \qquad \left| \begin{matrix} 1 & z^j\bar{z}^k & \bar{z} & z^2 & z\bar{z} & \bar{z}^2 \end{matrix} \right| = 0, \qquad \left| \begin{matrix} 1 & z & z^j\bar{z}^k & z^2 & z\bar{z} & \bar{z}^2 \end{matrix} \right| = 0.$$

9. Some further unsolved problems. Since it is useful to demarcate the frontiers of knowledge by pointing out what lies a little outside them as well as what is within, a few of the many questions may be mentioned which this paper falls short of answering. Besides the extension to two variables mentioned in the last section, and to an arbitrary number of variables, it is desirable that the whole theory should be developed from an exact, or small-sample, point of view rather than on the basis of the large-sample approximations used here. This however appears to be an extremely large enterprise. A simpler, but still quite difficult, problem is to modify the criteria obtained in paragraphs 6 and 7 so as to fit problems of economic experimentation, such as those of determination of maximum monopoly profit or minimum cost, in which the cost of each observation consists largely of the lost profit, or excess cost over the minimum, occasioned by the deviation from the value sought. In such a case the limitation of cost replaces the limitation of the total number of observations.

Another important problem is to take account of the inaccuracy of the preliminary information on which the design of the experiment is based, and to utilize the relations thus involved to design efficient sequences of experiments.

Determination of limits of error in terms of the maxima *over an interval* of the derivatives of $f(x)$ should be a fairly straightforward problem in analysis and have practical importance. With this are associated various problems dealing with maxima of functions having discontinuities in the first or higher derivatives at or near the maximum.

An important extension would deal with the case in which the maximum is estimated from a least-squares polynomial of degree three or more. This might be a connected with the difficult wider problem of deciding on the degree of a polynomial to be fitted in a particular case.

10. Summary. In determining the value ξ of x for which $f(x)$ is a maximum or minimum, a quadratic polynomial may be fitted to observations made for chosen values of x. The errors considered are of two kinds: sampling errors resulting from the inaccuracy in each observation, which diminish as the number of observations is increased, but increase if the values of x are chosen too close to the value sought; and biased errors resulting from the fact that $f(x)$ is not truly quadratic, which do not decrease when the number of observations increases with a fixed set of values of x, but do decrease when the deviations of x from the value sought are reduced. The biased errors may be separated into components corresponding to the third, fourth and higher powers of $x - \xi$ in the expansion of $f(x)$, and these components will ordinarily be of diminishing importance as we go on in the sequence. However it is possible to choose values of x making the cubic component zero and the quartic component at the same time a minimum. Such a set consists of only three values of x. These values may be further adjusted to minimize the expectation of the square of the total error in ξ, as far at least as the term of fourth order in the bias, by a proper balance between the sampling variance and the quartic bias. The values of x satisfying these conditions, measured from the true maximizing or minimizing value ξ, are the products of $[\sigma^2/(N\beta_4^2)]^{1/8}$ by the values u in the table below. Since the root will usually be extracted by logarithms, the common logarithms of the values are given. The first set are the most efficient when the frequencies must be equal. The second set is appropriate when the frequencies are made proportional to the quantities in the last column; in this case only about 72 per cent as many observations are required for any specified accuracy as when the frequencies must be equal. The approximate expected squared errors in the estimates of ξ in the two cases are given respectively by formulae (44) and (59). All these results are approximations of the kind appropriate to large numbers of observations.

Equal frequencies		*Adjustable frequencies*		
u	$\log_{10} u$	u	$\log_{10} u$	Frequency
$-.6128$	$-.21267$	$-.2110$	$-.67572$	$46\ 051\ N$
$.8695$	$-.06071$	$.7877$	$-.10364$	$.50\ 000\ N$
2.0751	$.31704$	2.1520	$.33284$	$.03\ 349\ N$

The signs of u should be reversed if $\beta_2\beta_4(\beta_2\beta_6 - 4\beta_3\beta_4) > 0$. Here β_k is the coefficient of $(x - \xi)^k$ in the expansion of $f(x)$, and σ^2 is the error variance of an individual observation. For designing an efficient experiment it is necessary

to have some knowledge of these quantities. It may be gained from preliminary experiments of smaller scale.

A suitable preliminary experiment, where knowledge of the function is extremely scanty, might consist of a fixed small number, greater than one, of observations on $f(x)$ corresponding to each of a set of six or more values of x in arithmetic progression covering an interval that includes the value ξ sought, and selected with a view to getting ξ in the center of it as nearly as possible. A polynomial of the fifth degree at least should be fitted by least squares, in which process all the quantities desired for the design of the later, larger experiment can be estimated, together with their accuracies. Since the values of x are taken in arithmetic progression, the fitting can be carried out with extreme ease by the method of orthogonal polynomials.

Numerous subsidiary questions promise to have both practical importance and mathematical interest.

REFERENCES

[1] Felix Bernstein, *Hartmann and Bauer Handbuch der Erblichkeitslehre*, Vol. 1, Berlin, 1928

[2] S. O Rice, "The distribution of the maximum of a random curve," *Am Jour of Math.*, Vol 61 (1939), pp. 409-416.

[3] G. W. Elmen, "Magnetic alloys of iron, nickel and cobalt in communication circuits," *Elec. Eng.*, Vol. 54 (1935), pp. 1292-1299.

[4] E. E. Schumacher and A G. Souden, "Some alloys of copper and iron," *Metals and Alloys*, Vol. 7 (1936), pp 95-101

[5] P. C. Mahalanobis, "A sample survey of the acreage under jute in Bengal, with discussion on planning of experiments," *Proc. 2nd Ind. Stat. Conf.*, Calcutta, Statistical Publishing Soc. (1940).

[6] H. G. Romig, *Allowable average in sampling inspection*, Columbia Univ. thesis, privately printed, 1939

[7] J. V Uspensky, *Introduction to mathematical probability*, New York, McGraw-Hill, 1937.

[8] R. A. Fisher, *The design of experiments*, London and Edinburgh, Oliver and Boyd, 1935, Chap. 6.

ON A STATISTICAL PROBLEM ARISING IN ROUTINE ANALYSES AND IN SAMPLING INSPECTIONS OF MASS PRODUCTION

By J. Neyman

University of California, Berkeley, Calif.

CONTENTS

	Page
1. Introduction	46
2. Statistical hypothesis H to be tested	48
3. General problem of similar regions	48
4. Regions similar to the sample space with regard to σ, ξ_1, ξ_2, $\cdots$, ξ_N	55
5. The set of hypotheses alternative to H	61
6. The best critical region for testing H against a particular alternative	65
7. A critical region of an unbiased type	66
8. Methods of determining ζ_1	70
9. References	76

1. **Introduction.** The words "routine analyses" are used to denote the analyses performed by laboratories, frequently attached to industrial plants, and distinguished by the following characteristics: (1) All the analyses or measurements are of the same kind, for example, are designed to measure the sugar content in beets or to determine the coordinate of a star. (2) The analyses are carried out day after day using the same methods and the same instruments. (3) While all the analyses are of the same kind, the quantity measured varies from time to time and each such quantity is measured repeatedly n times, where n represents some small number, 2, 3, 4, 5.

As an illustration we may consider the routine analyses of sugar beets performed in the process of selection and breeding. A small section is cut out of each of a great number of sugar beets expected to be suitable for further breeding. It is crushed and its juice extracted to determine ξ, the sugar content of each particular beet. From the juice available from each beet n samples are taken and a determination of the sugar content is made from each. Thus, if ξ_i represents the sugar content of the section from the ith beet and there are N beets, the laboratory will have to make nN analyses with their results $x_{i,1}$, $x_{i,2}$, $\cdots$, $x_{i,n}$, representing the measurements of the same quantity ξ_i. Obviously the sugar content ξ_i referring to the ith beet need have no relation to that of any other jth beet.

An essential point in the above description is that the number of measurements referring to the same quantity ξ_i is usually very small. For example, the quantitative analyses of urine in certain clinics are performed only twice for each patient, so that $n = 2$. Frequently, various practical considerations make

it impossible to increase this number n of analyses intended to measure the same quantity ξ_i.

The smallness of n introduces difficulties in estiminating ξ_i. It is usual to consider $x_{i,1}, x_{i,2}, \cdots, x_{i,n}$ as independent variables, varying normally about ξ_i with an unknown standard error σ_i. If they have to be used to estimate ξ_i, then the confidence interval [1][1] for ξ_i will be determined by the familiar formula

$$x_{i.} - s_i t_\alpha(n) \leq \xi_i \leq x_{i.} + s_i t_\alpha(n), \tag{1}$$

where $x_{i.}$ denotes the mean of the x_{ij},

$$s_i^2 = \sum_{j=1}^{n} (x_{ij} - x_{i.})^2/n(n-1) \tag{2}$$

and $t_\alpha(n)$ is Fisher's t corresponding to the number of degrees of freedom $n - 1$ and to the chosen confidence coefficient α. It is known [2] that if the estimate of ξ_i is based only on its direct measurements $x_{i,1}, x_{i,2}, \cdots, x_{i,n}$, then the confidence interval (1) can not be made any smaller; in fact, formula (1) gives the shortest unbiased confidence interval for ξ_i. But if we try to substitute appropriate numbers in (1) we get disconcerting results. Namely, if $n = 2$ and $\alpha = .99$, then $t_\alpha(n) = 63\ 657$. If n is increased, the value of $t_\alpha(n)$ decreases rapidly but for $n = 5$ it is still very considerable, $t_\alpha(5) = 4.604$, and consequently the numerical confidence interval determined by (1) is frequently so broad that it is devoid of practical value.

The general conclusion is that, if n cannot be increased, satisfactory estimates of ξ_i can only be obtained when they are based on something else in addition to the direct measurements $x_{i,1}, x_{i,2}, \cdots, x_{i,n}$. This point was first noticed by "Student" [3]. His method of avoiding the difficulty consists in assuming that the accuracy of measurements performed in the same laboratory is constant in time, so that $\sigma_1 = \sigma_i = \cdots = \sigma_N = \sigma$. If this is true, then $s_0^2 = \Sigma s_i^2/N$ will be an unbiased estimate of the variance of $x_{i,j}$, based on $N(n-1)$ degrees of freedom. If the past experience of the laboratory is of any size, as measured by N, then the product $N(n-1)$ will be of considerable size and the confidence interval for ξ_i

$$x_{i.} - s_0 t_\alpha(N(n-1)+1) \leq \xi_i \leq x_{i.} + s_0 t_\alpha(N(n-1)+1) \tag{3}$$

will be much more satisfactory than (1).

The problem which arises is whether we are entitled to assume that $\sigma_1 = \sigma_2 = \cdots = \sigma_N$. The first study of this problem seems to have been made by Przyborowski [4] in a paper written in Polish. His findings, subsequently reported [5] in English, show that, at least in certain cases, the accuracy of routine analyses is quite difficult to keep constant. If it is not constant, then the relative frequency of the cases where formula (3) gives correct statements about ξ_i will generally be different from the expected α.

[1] Figures in square brackets refer to the literature quoted at the end of the paper.

The procedure employed by Przyborowski to test whether $\sigma_1 = \sigma_2 = \cdots = \sigma_N$ consisted in considering the quantities $v_i = (n-1)s_i^2$ and applying the χ^2 test to see whether they follow the same χ^2 distribution with $n-1$ degrees of freedom

$$p(v) = cv^{\frac{1}{2}n-\lambda_i}e^{-\frac{1}{2}v/\sigma^2} \tag{4}$$

with an unknown σ.

Just this point is to be the main subject of this paper. The χ^2 test was devised by Karl Pearson with no particular set of alternative hypotheses in view. As a result we may expect that in many cases other tests may be devised which would be more powerful. A number of such cases are already on record [6], [7], [8].

2. Statistical hypothesis H to be tested. We shall consider the case where we can observe the particular values of Nn random variables $x_{i,j}$, $i = 1, 2, \cdots, N$; $j = 1, 2, \cdots, n$, and we know that $x_{i,j}$ is independent of x_{kl} for $i \neq k$ and that

$$p(x_{i,1}, x_{i,2}, \cdots x_{i,n}) = \left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{j=1}^{n}(x_{i,j}-\xi_i)^2/\sigma_i^2} \tag{5}$$

with unknown values of ξ_i and $\sigma_i > 0$. The hypothesis H to be tested is that $\sigma_1 = \sigma_2 = \cdots = \sigma_N = \sigma$ without specifying, however, the actual value of σ. It will be noticed that this hypothesis has already been treated by a number of authors [9]–[17]. The need for considering it again arises from the fact that previously it was tested against the set of alternatives presuming that the σ_1, $\sigma_2, \cdots \sigma_N$, were positive constants having any values whatsoever. It seems to the author that, in the present case, the set of alternatives should be different. This will be explained in the next section. It follows that while the hypothesis tested is the same as in the papers quoted above, the problem of testing it is quite different.

Let us denote by E the whole set of Nn observable variables. If H is true then their elementary probability law will be

$$p(E \mid H) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{Nn} e^{-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{n}(x_{i,j}-\xi_i)^2/\sigma^2} \tag{6}$$

3. General problem of similar regions. The development of the test will follow the general lines explained elsewhere [18], [19], [20]. Denoting by W the Nn dimensional space of the $x_{i,j}$'s, we want to determine a region w in W having the following properties: (a) if the hypothesis tested is true then the probability of E falling in w shall have some fixed value chosen in advance, e.g., $\epsilon = .05$ or $\epsilon = .01$. This probability is known as the probability of an error of the first kind. (b) If H is not true then the probability of E falling in w as determined by one of the alternative hypotheses (that we assume likely to be true when H is false) shall be as large as possible in a sense that requires further explanation.

The probability with which this condition is concerned is a complement of the probability of an error of the second kind. Once the region w is chosen it will be used to test H in this way. if E falls within w, then H will be rejected

In the present section we shall deal only with ways of satisfying condition (a). The problem is similar to the one recently described by Hotelling [21]. The difficulty is that, if H is true, the probability law of E is given by (6) and contains $N + 1$ unspecified parameters, "nuisance" parameters as Hotelling very appropriately calls them. If we take just any region w then it is most likely that the probability of E falling in it will vary with different values of $\sigma, \xi_1, \cdots, \xi_N$. As a matter of fact, if we want the test to be absolutely most powerful, or at least relatively so, we must determine not just one single region satisfying (a) but actually *all* such regions or some broad family of them. From these we shall then select one which seems most satisfactory from the point of view of (b).

Systematic methods of determining regions of the above kind have already been considered [18], [20], [2] In these publications they are called "similar" to the sample space W. The reason for this term is that the whole space W does possess the required properties with $\epsilon = 1$ In fact, whatever be the values of the nuisance parameters, $\sigma, \xi_1, \cdots, \xi_N$, the probability of E falling within W, as calculated from (6), is perfectly determined and equals 1. Our problem is to find a region w, part of W, with similar properties for $0 \leq \epsilon < 1$. However, in many cases no such regions exist [22].

The general methods in the above publications are applicable in the present case. However, a recent paper by Cramér and Wold [23] allows a slight improvement in presenting the matter. As this is a little involved, it seems desirable to take up the whole problem and present it anew.

Consider then the general case where the probability law of some m observable variables $y_1, y_2, \cdots, y_m$, say $p(E \mid \theta_1, \cdots, \theta_s)$, as specified by the hypothesis tested, depends on s nuisance parameters $\theta_1, \theta_2, \cdots, \theta_s$. Our problem will consist of determining the necessary and sufficient conditions for a region w to be similar to the sample space with respect to all these parameters. We shall assume that the probability law $p(E \mid \theta_1, \cdots, \theta_s)$ satisfies certain limiting conditions

Let

$$\varphi_i = \frac{\partial \log p}{\partial \theta_i} \tag{7}$$

$$\varphi_{ij} = \frac{\partial^2 \log p}{\partial \theta_i \partial \theta_j} \tag{8}$$

Assume that for all values of i and $j = 1, 2, \cdots, s$

$$\varphi_{ij} = A_{ij} + \sum_{k=1}^{s} B_{ij,k} \varphi_k \tag{9}$$

where the coefficients A_{ij} and $B_{ij,k}$ are independent of the observable variables E. Assume also that the probability law $p(E \mid \theta_1, \cdots, \theta_s)$ permits indefinite

differentiation under the sign of the integral taken over any fixed region w in W. It is easy to check that the probability law (6) satisfies all of these conditions.

In order to find the necessary conditions for the region w to be similar to W with respect to $\theta_1, \theta_2, \cdots, \theta_s$, assume that w is actually similar and that, consequently,

$$P\{E \epsilon w \mid \theta_1, \cdots, \theta_s\} = \int \cdots \int_w p(E \mid \theta_1, \cdots \theta_s)\, dy_1 \cdots dy_m = \epsilon \tag{10}$$

for all possible values of $\theta_1, \theta_2, \cdots, \theta_s$. It follows that the derivatives of all orders with respect to $\theta_1, \theta_2, \cdots, \theta_s$ taken from the left side of (10) must be identically equal to zero. But we have

$$\begin{aligned} &\frac{\partial}{\partial \theta_i} \int \cdots \int_w p(E \mid \theta_1, \cdots, \theta_s)\, dy_1 \cdots dy_m \\ &\qquad = \int \cdots \int_w \frac{\partial}{\partial \theta_i} p(E \mid \theta_1, \cdots, \theta_s)\, dy_1 \cdots dy_m \\ &\qquad = \int \cdots \int_w \varphi_i p(E \mid \theta_1, \cdots, \theta_s)\, dy_1 \cdots dy_m = 0 \end{aligned} \tag{11}$$

for $i = 1, 2, \cdots, s$. Similarly, using (9)

$$\begin{aligned} &\frac{\partial^2}{\partial \theta_i \partial \theta_j} \int \cdots \int p(E \mid \theta_1, \cdots \theta_s)\, dy_1 \cdots dy_m \\ &= \int \cdots \int_w \left(\varphi_i \varphi_j + A_{ij} + \sum_{k=1}^{s} B_{i,j,k} \varphi_k \right) p(E \mid \theta_1, \cdots, \theta_s)\, dy_1 \cdots dy_m = 0. \end{aligned} \tag{12}$$

Using (10) and (11), the last identity will be reduced to

$$\frac{1}{\epsilon} \int \cdots \int_w \varphi_i \varphi_j p(E \mid \theta_1, \cdots \theta_s)\, dy_1 \cdots dy_m = -A_{i,j} \quad \text{for } i, j = 1, 2, \cdots, s \tag{13}$$

where the right side does not depend on the particular region w, provided that w is similar to the sample space. Considering the identities (11) and (13) which were obtained by differentiating (10) twice, we may guess what will happen if we differentiate (13) again and again. We may assume, in fact, that, whatever be the non-negative integers $k_1, k_2, \cdots, k_s$, we shall obtain

$$\frac{1}{\epsilon} \int \cdots \int_w \prod_{i=1}^{s} \varphi_i^{k_i} p(E \mid \theta_1, \cdots, \theta_s)\, dy_1 \cdots dy_m = M(k_1, k_2, \cdots, k_s), \tag{14}$$

where $M(k_1, \cdots, k_s)$ is independent of the particular region w, provided that w is similar to the sample space with respect to all of the θ's. Assume that this is found for all k's such that $\sum_{i=1}^{s} k_i \leq K$; also assume that the sum of the k's in (14) is exactly K. Differentiating with respect to θ_j, we obtain

$$\begin{aligned} &\frac{1}{\epsilon} \int \cdots \int_w \left\{ \varphi_j \prod_{i=1}^{s} \varphi_i^{k_i} + \prod_{i=1}^{s} \varphi_i^{k_i} \sum_{i=1}^{s} \varphi_i^{-1} \varphi_{i,j} \right\} p(E \mid \theta_1, \cdots, \theta_s)\, dy_1 \cdots dy_m \\ &\qquad = \frac{\partial}{\partial \theta_j} M(k_1, \cdots, k_s). \end{aligned} \tag{15}$$

Because of the particular form of $\varphi_{i,j}$, the second expression in the curly brackets under the integral is a polynomial in the φ's of order not exceeding K. According to the assumption made, this expression multiplied by $p(E \mid \theta_1, \cdots, \theta_s)/\epsilon$ and integrated over w gives a result which is independent of w. As the right side of (15) is also independent of w, we conclude that

$$\frac{1}{\epsilon}\int \cdots \int_w \varphi_j \prod_{i=1}^{s} \varphi_i^{k_i}\, p(E \mid \theta_1, \cdots, \theta_s)\, dy_1 \cdots dy_m \equiv M(k_1, \cdots, k_j + 1, \cdots, k_s) \tag{16}$$

is also independent of the particular similar region chosen. We have seen that (14) is true for $K \leq 2$ and that if it is true for K it is true for $K + 1$, that is, it is true in general.

We may now sum up our findings: if w is a region similar to the sample space with respect to all of the θ's and if ϵ denotes the value of the integral (10), then, whatever be the non-negative integers $k_1, k_2, \cdots, k_s$, the value of the integral on the left side of (14) is independent of the particular region w chosen.

As the whole sample space W is also "similar" with $\epsilon = 1$, it must satisfy this identity. This allows us to determine the M's, namely

$$\int \cdots \int_W \prod_{i=1}^{s} \varphi_i^{k_i}\, p(E \mid \theta_1, \cdots, \theta_s)\, dy_1 \cdots dy_m \equiv M(k_1, \cdots, k_s). \tag{17}$$

It is obvious that the necessary condition above is also sufficient. If a region w is such that (14) holds for all systems of non-negative integers then all the derivatives of (10) must be identically zero; thus the left side of (10) is independent of $\theta_1, \theta_2, \cdots, \theta_s$.

It will be useful to interpret the above conditions as follows. We start by noticing that the left side of (17) represents the product moment of some specified order of the $\varphi_1, \varphi_2, \cdots, \varphi_s$ considered as random variables. We shall call it the absolute product moment. We will now interpret the left side of (14) as a product moment also. For this purpose we shall define a new elementary probability law of the y's to be denoted by $p(E \mid w, \theta_1, \cdots, \theta_s)$ and described as the relative probability law given w. We shall write it as

$$p(E \mid w, \theta_1, \cdots, \theta_s) = \frac{1}{\epsilon}\, p(E \mid \theta_1, \cdots, \theta_s) \tag{18}$$

for all of the points E included in w and

$$p(E \mid w, \theta_1, \cdots, \theta_s) = 0 \tag{19}$$

for all other points. With this definition the left side of (14) appears to be the expectation of the product $\varphi_1^{k_1} \cdots \varphi_s^{k_s}$ calculated from the relative probability law of the y's given w. We will call it the relative product moment given w. The final result can now be stated as follows:

For a region w to be similar to the sample space with respect to $\theta_1, \theta_2, \cdots, \theta_s$ it is necessary and sufficient that all the relative moments and product moments

of $\varphi_1, \varphi_2, \cdots, \varphi_s$ shall equal the corresponding absolute moments and product moments.

In order to make the method of constructing similar regions according to the above conditions clear we recall the procedure involved in the calculation of the probability laws of any given set of random variables.

Assume then that the elementary probability law of the original variables is given. Fix some values of the parameters $\theta_1, \theta_2, \cdots, \theta_k$, denote the resulting probability law by $p(E)$, and consider the problem of finding the elementary probability law of $\varphi_1, \varphi_2, \cdots, \varphi_s$ considered as functions of the y's. We shall assume that none of the φ's can be expressed as a function of the others not involving the y's explicitly so that the matrix

$$\begin{pmatrix} \frac{\partial \varphi_1}{\partial y_1} & \frac{\partial \varphi_1}{\partial y_2} & \cdots & \frac{\partial \varphi_1}{\partial y_m} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \varphi_s}{\partial y_1} & \frac{\partial \varphi_s}{\partial y_2} & \cdots & \frac{\partial \varphi_s}{\partial y_m} \end{pmatrix} \tag{20}$$

is non-singular. In these circumstances it is possible to select $m - s$ functions of the y's say $\psi_{s+1}, \psi_{s+2}, \cdots, \psi_m$ which have continuous second derivatives such that the formulae

$$\begin{aligned} z_i &= \varphi_i \qquad && i = 1, 2, \cdots, s \\ z_j &= \psi_j \qquad && j = s+1, \cdots, m \end{aligned} \tag{21}$$

determine a one-to-one transformation of the space W of the y's into the space W' of the z's. If w denotes any region in W then it will be transformed into a perfectly determined region w' in W'. If E' denotes a point in W' then the probability of E' falling in w' will be identical with that of E falling in w. Thus

$$P\{E' \epsilon w'\} \equiv P\{E \epsilon w\} = \int \cdots \int_w p(E)\, dy_1 \cdots dy_m. \tag{22}$$

Letting J be the Jacobian of the y's with respect to the z's in the transformation (21) and using the known formulae for transforming multiple integrals, we have

$$P\{E' \epsilon w'\} = \int \cdots \int_{w'} p(E)\Big]_{E'} |J|\, dz_1 \cdots dz_m, \tag{23}$$

where $p(E)]_{E'}$ denotes the result of substituting the expressions for the y's in terms of the z's as obtained from (21) into $p(E)$. It follows that, whatever be the region w' in W', the probability of E''s falling in it is obtained by integrating the function $p(E)]_{E'} |J|$ over w'. But this means, according to the usual definition, that the product $p(E)]_{E'} |J|$ is the elementary probability law of the z's. Denoting it by $p(E') = p(z_1, \cdots, z_m)$ we have

$$p(E') = p(E)]_{E'} |J|. \tag{24}$$

Now, to obtain the joint probability law of $\varphi_1, \varphi_2, \cdots, \varphi_s$ or that of $z_1, z_2, \cdots, z_s$ we must integrate $p(E')$ for all the other z's between their extreme limits, formally between $-\infty$ and $+\infty$ for each of the variables concerned,

$$p(\varphi_1, \cdots, \varphi_s) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p(E')\, dz_{s+1} \cdots dz_m. \tag{25}$$

This procedure will be applied when calculating the absolute probability law of the φ's and also the relative one given w. The only difference will be that in the latter case we shall have to start with (18) and (19) instead of the original probability law. The space W' and the transformation (21) will be the same in both cases. It is important to be clear about the difference between the two cases. This is connected with the difference between $p(E \mid \theta_1, \cdots, \theta_s)$ and $p(E \mid w, \theta_1, \cdots, \theta_s)$ of (18) and (19) The latter is proportional to the former at any point E within the region w but is zero outside of w. As mentioned above, the integrations for $z_{s+1}, z_{s+2}, \cdots, z_m$ in (25) should extend formally from $-\infty$ to $+\infty$ for each variable. However, the probability law $p(E')$ may equal zero within certain parts of this range. Fixing any system of values $z_i = \varphi_i$, for $i = 1, 2, \cdots, s$, is equivalent to fixing a hypersurface in the space W and considering the intersection of planes $z_i = \text{constant}$ in the space W'. Denote them by $W(\varphi)$ and $W'(\varphi)$, respectively. If we shift the point E or E' along $W(\varphi)$ or $W'(\varphi)$ respectively, the variables $z_j = \psi_j$, for $j = s + 1, s + 2, \cdots, m$ will assume a certain set $S(\varphi)$ of systems of values. When calculating the absolute probability law of $\varphi_1, \cdots, \varphi_s$ this set $S(\varphi)$ will be the real region of integration in (25); outside of it the function under the integral sign will be zero. On the other hand, when calculating the relative probability law of $\varphi_1, \cdots, \varphi_s$ given w, the function under the integral (25) is zero as soon as the point E moves outside of the region w. Denote by $w(\varphi)$ that part of $W(\varphi)$ which is included in w and by $w'(\varphi)$ the corresponding part of $W'(\varphi)$. So, the absolute and the relative, given w, probability laws of $\varphi_1, \cdots, \varphi_s$ can be obtained by using the formulae

$$p(\varphi_1, \cdots, \varphi_s) = \int \cdots \int_{W'(\varphi)} p(E')\, dz_{s+1} \cdots dz_m \tag{26}$$

$$p(\varphi_1, \cdots, \varphi_s \mid w) = \frac{1}{\epsilon} \int \cdots \int_{w'(\varphi)} p(E')\, dz_{s+1} \cdots dz_m. \tag{27}$$

Now the method of constructing regions similar to W with respect to $\theta_1, \theta_2, \cdots, \theta_s$ is clear: to construct any such region it is necessary and sufficient to select for each of all possible systems of values of $\varphi_1, \varphi_2, \cdots, \varphi_s$ a part $w(\varphi)$ of the hypersurface $W(\varphi)$ and to combine all these parts. The selection of $w(\varphi)$ is arbitrary save for the restriction that the probability law (27) have all its moments equal to those of (26), identically in the θ's. This last condition will

certainly be satisfied if $w(\varphi)$ is so selected that for almost all systems of values of $\varphi_1, \varphi_2, \cdots, \varphi_s$

$$p(\varphi_1, \cdots, \varphi_s \mid w) \equiv p(\varphi_1, \cdots, \varphi_s) \tag{28}$$

for all values of the θ's.

By selecting $w(\varphi)$ in all possible ways that satisfy (28) we obtain an infinity of regions similar to W with respect to $\theta_1, \theta_2, \cdots, \theta_s$. They form a family which we shall denote by $F(\epsilon)$. However, it is known that in general all the moments of $p(\varphi_1, \cdots, \varphi_s \mid w)$ and $p(\varphi_1, \cdots, \varphi_s)$ may be identical without the two probability laws being equal almost everywhere. In such cases, the family $F(\epsilon)$ will not exhaust all the similar regions. It is important to be able to state whether or not $F(\epsilon)$ contains all the similar regions. To ascertain this we may use the conditions of Cramér and Wold [23] which are sufficient for the determinateness of the problem of moments, that is, for the uniqueness of a function having a given set of moments.

Let

$$\mu_\nu = M(\nu, 0, 0, \cdots, 0) + M(0, \nu, 0, \cdots, 0) + \cdots + M(0, 0, \cdots, 0, \nu). \tag{29}$$

With this notation the conditions of Cramér and Wold can be stated as follows: If any two probability laws, e.g., the probability laws $p(\varphi_1, \cdots, \varphi_s \mid w)$ and $p(\varphi_1, \cdots, \varphi_s)$, have all their moments and all their product moments identical and if the series

$$\sum_\nu \mu_{2\nu}^{-1/2\nu} \tag{30}$$

is divergent, then

$$p(\varphi_1, \cdots, \varphi_s \mid w) \equiv p(\varphi_1, \cdots, \varphi_s) \tag{31}$$

almost everywhere.

Therefore, to know whether the family $F(\epsilon)$ defined above exhausts all the regions similar to W, we must calculate the even moments of all the φ_i and see whether the series (30) depending on these moments is divergent. If it is, there is no similar region besides the family $F(\epsilon)$. Otherwise, there may be some others. These others will be constructed by selecting $w(\varphi)$'s such that the integral (27) equals any other probability law having the same moments as (26). In such cases, a region w selected, in one way or another, from the family $F(\epsilon)$ as the best from the point of view of controlling errors of the second kind will only be the relative best.

It should be mentioned that whether we can always, under the conditions considered, select a $w(\varphi)$ on any $W(\varphi)$ that satisfies the identity (28) has not yet been proved. However, it seems plausible that the differential equations (9) imply the existence of a sufficient set of statistics for $\theta_1, \theta_2, \cdots, \theta_s$. If this is so, the possibility of satisfying (28) is guaranteed (see [2], p. 366).

4. **Regions similar to the sample space with respect to $\sigma, \xi_1, \xi_2, \cdots, \xi_N$.** We may now return to the original problem and apply our theory to the probability law (6). We wish to construct the most general regions similar to the sample space with respect to the nuisance parameters $\sigma, \xi_1, \cdots, \xi_N$ unspecified by the hypothesis tested. We let

$$\varphi_\sigma = \frac{\partial \log p}{\partial \sigma} = -\frac{Nn}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{N}\sum_{j=1}^{n}(x_{ij} - \xi_i)^2, \tag{32}$$

$$\varphi_i = \frac{\partial \log p}{\partial \xi_i} = \frac{n(x_{i\cdot} - \xi_i)}{\sigma^2} \quad \text{with} \quad x_{i\cdot} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}. \tag{33}$$

Then

$$\begin{aligned} \frac{\partial \varphi_\sigma}{\partial \sigma} &= -\frac{3}{\sigma}\varphi_\sigma - \frac{2Nn}{\sigma^2} \\ \frac{\partial \varphi_\sigma}{\partial \xi_i} &= -2\sigma\varphi_i \\ \frac{\partial \varphi_i}{\partial \xi_i} &= -\frac{n}{\sigma^2} \\ \frac{\partial \varphi_i}{\partial \xi_j} &= 0, \qquad i \neq j \end{aligned} \tag{34}$$

and we see that the probability law (6) satisfies the differential equations (9).

Now the hypersurfaces $W(\varphi)$ of the theory are the intersections of the hypersurfaces

$$\varphi_\sigma = \text{constant and} \qquad \varphi_i = \text{constant}, \qquad \text{for } i = 1, 2, \cdots, N. \tag{35}$$

The latter equations are clearly equivalent to

$$x_{i\cdot} = \text{constant}. \tag{36}$$

As to the former, we notice the identity

$$\sum_{i=1}^{N}\sum_{j=1}^{n}(x_{i,j} - \xi_i)^2 = n\sum_{i=1}^{N}(S_i^2 + (x_{i\cdot} - \xi_i)^2) = \chi^2, \text{ (say)} \tag{37}$$

where $nS_i^2 = \sum_{j=1}^{n}(x_{i,j} - x_i)^2$. Therefore, $W(\varphi)$ denotes the intersection of the hypersurfaces (36) with, say,

$$T_1 = \sum_{i=1}^{N} S_i^2 = \text{constant}. \tag{38}$$

If we succeed in selecting from each hypersurface $W(\varphi)$ a part $w(\varphi)$ satisfying condition (28) identically then the sum of all such regions $w(\varphi)$ will form a region w similar to W with respect to all the unspecified parameters and belonging to the family $F(\epsilon)$. Before proceeding to this stage of the solution, let us see whether the family $F(\epsilon)$ exhausts all of the similar regions.

For this purpose notice first that instead of considering whether there is but one probability law with moments equal to those of φ_s and the φ_s's, it is sufficient to concern ourselves with the moments of χ^2 and x_i. In fact, all the φ's are functions of these variables and the problem of uniqueness of the distribution must have the same answer in both cases. The 2νth absolute moment of χ^2 as calculated from (6) equals

$$(2\sigma^2)^{2\nu}\Gamma(\tfrac{1}{2}Nn + 2\nu)\,[\Gamma(\tfrac{1}{2}Nn). \tag{39}$$

The same order moment of x_i is

$$\sigma^{2\nu}(2\nu)!/(2n)^{\nu}\nu!. \tag{40}$$

Thus, the quantity denoted by $\mu_{2\nu}$ in the theory becomes

$$\mu_{2\nu} = \frac{(2\sigma^2)^{2\nu}\Gamma(\tfrac{1}{2}Nn + 2\nu)}{\Gamma(\tfrac{1}{2}Nn)} + N\left(\frac{\sigma^2}{n}\right)^{\nu}\frac{(2\nu)!}{2^{\nu}\nu!}. \tag{41}$$

We are interested in whether or not the series (30) is divergent. Since $\mu_{2\nu}$ satisfies the inequality

$$\mu_{2\nu} < a^{2\nu}\Gamma(b + 2\nu) = C_{2\nu}^{-2\nu}, \qquad \text{(say)} \tag{42}$$

with $a = 2\sigma^2 + N$ and $2b = Nn$, if we prove that the series $\Sigma C_{2\nu}$ diverges, then (30) also diverges. To settle this conveniently we apply Stirling's formula to $\Gamma(b + 2\nu)$ and find that, as $\nu \to \infty$, the ratio $C_{2\nu}/\nu^{-1}$ tends to a finite limit. As the series $\Sigma\nu^{-1}$ is divergent, so is the series $\Sigma C_{2\nu}$ and thus the series $\Sigma\mu_{2\nu}^{1/2\nu}$ is divergent. Therefore, there is but one probability law with moments identical to those of χ^2 and the x_i's and so the family $F(\epsilon)$ contains all the regions similar to the sample space with respect to $\sigma, \xi_1, \cdots, \xi_N$.

It may now be interesting to go into some details of the effective construction of any region similar to W with respect to $\sigma, \xi_1, \cdots, \xi_N$. For this purpose it is convenient to go back and express the identity (28), that the regions $w(\varphi)$ must satisfy, in terms of the relative probability law of $z_{s+1}, z_{s+2}, \cdots, z_m$ given $\varphi_1, \varphi_2, \cdots, \varphi_s$. This is denoted by $p(z_{s+1}, z_{s+2}, \cdots, z_m \mid \varphi_1, \cdots, \varphi_s)$ and defined for every system of values of the φ's for which $p(\varphi_1, \varphi_2, \cdots, \varphi_s) \neq 0$ as follows:

$$p(z_{s+1}, z_{s+2}, \cdots, z_m \mid \varphi_1, \varphi_2, \cdots, \varphi_s) = p(\varphi_1, \cdots, \varphi_s, z_{s+1}, \cdots, z_m)/p(\varphi_1, \cdots, \varphi_s). \tag{43}$$

Using (26), (27), and (43), the identity (28) can be rewritten in the following form

$$\int \cdots \int_{w'(\varphi)} p(z_{s+1}, \cdots, z_m \mid \varphi_1, \cdots, \varphi_s)\, dz_{s+1} \cdots dz_m = \epsilon. \tag{44}$$

The function under this integral is the relative elementary probability law of $z_{s+1}, z_{s+2}, \cdots, z_m$ and it is integrated over the region $w'(\varphi)$. Therefore, the left side of (44) is nothing but the relative probability of the point E' falling in

$w'(\varphi)$ given that the first s of its coordinates have the fixed values $\varphi_1, \varphi_2, \cdots, \varphi_s$. In other words, and owing to the one-to-one correspondance between the spaces W and W', we have

$$P\{E' \epsilon w'(\varphi) \mid E' \epsilon W'(\varphi)\} = P\{E \epsilon w(\varphi) \mid E \epsilon W(\varphi)\} \equiv \epsilon. \tag{45}$$

Now the general method of determining similar regions may be stated as follows:

1. Choose any system of variables $z_{s+1}, z_{s+2}, \cdots, z_m$ such that their values determine uniquely the position of the point E' on any fixed hypersurface $W'(\varphi)$ These z's considered as functions of the y's should be continuously differentiable twice.

2. Find the relative probability law of the z's given the φ's. This must be done for every possible set of values of the φ's.

3. In the space of $z_{s+1}, z_{s+2}, \cdots, z_m$ consider regions which satisfy the equality (44) identically in the θ's. Any such region could be taken to form a part of w', the region similar to the sample space, which we are trying to construct. If the assumption that the differential equations (9) imply the existence of a sufficient system of statistics for $\theta_1, \theta_2, \cdots, \theta_s$ is true, then (see [2], p. 366) the probability law $p(z_{s+1}, z_{s+2}, \cdots, z_m \mid \varphi_1, \cdots, \varphi_s)$ will be independent of the θ's and there will be an infinity of regions satisfying (44)

Obviously, instead of dealing directly with $\varphi_1, \varphi_2, \cdots, \varphi_s$ as described above, we may select any system of statistics $T_1, T_2, \cdots, T_s$ such that the system of equations $T_i = \text{constant}$ is equivalent to $\varphi_i = \text{constant}$, for $i = 1, 2, \cdots, s$.

Returning to the particular problem of similar regions with respect to σ, $\xi_1, \cdots, \xi_N$, we notice that instead of the φ's we may consider

$$T_1 = \sum_{i=1}^{N} S_i^2 \quad \text{and} \quad T_{i+1} = x_i. \qquad \text{for } i = 1, 2, \cdots, N. \tag{46}$$

Now we wish to select a convenient system of variables, denoted by z_{s+i}'s in the theory above, to determine the position of the point E' on any hypersurface $W'(\varphi)$ where all the functions (46) have fixed values. Obviously there is no unique choice and we shall use what we find convenient. But notice that the total number of these variables should be, in our case, $Nn - N - 1$. The following system may be suggested.

If the sum ΣS_i^2 has a fixed value T_1 then none of the S_i^2 can exceed T_1. Write

$$\begin{aligned} S_i^2 &= u_i T_1 \\ S_N^2 &= \left(1 - \sum_{i=1}^{N-1} u_i\right) T_1 \qquad \text{for } i = 1, 2, \cdots, N-1 \end{aligned} \tag{47}$$

and consider $u_1, u_2, \cdots, u_{N-1}$ as belonging to the system of variables sought. The region of their variation is determined by the inequalities

$$0 \leq u_i \quad \text{and} \quad \sum_{i=1}^{N-1} u_i \leq 1 \tag{48}$$

If the u's are fixed then they, together with the value of T_1, determine the values of $S_1, S_2, \cdots, S_N$. As the values of $x_i = T_{i+1}$ are already fixed, we have to solve the problem of choosing for each $i = 1, 2, \cdots, N$ a system of $n - 2$ variables, say $z_{i,1}, z_{i,2}, \cdots, z_{i,n-2}$, which with x_i. and S_i will completely determine the values of $x_{i,1}, x_{i,2}, \cdots, x_{i,n}$. However, this will only have to be done if $n > 2$. Following the now familiar method (see, for example, [5], pp. 33–43), we may determine the $z_{i,j}$ in two consecutive steps. First write

$$
\begin{aligned}
x_{i,1} &= x_{i.} + \sqrt{\frac{1}{1\cdot 2}}\, v_{i,1} + \sqrt{\frac{1}{2\cdot 3}}\, v_{i,2} + \cdots + \sqrt{\frac{1}{(n-1)n}}\, v_{i,n-1} \\
x_{i,2} &= x_{i.} - \sqrt{\frac{1}{1\cdot 2}}\, v_{i,1} + \sqrt{\frac{1}{2\cdot 3}}\, v_{i,2} + \cdots + \sqrt{\frac{1}{(n-1)n}}\, v_{i,n-1} \\
x_{i,3} &= x_{i.} - 2\sqrt{\frac{1}{2\cdot 3}}\, v_{i,2} + \cdots + \sqrt{\frac{1}{(n-1)n}}\, v_{i,n-1} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
x_{i,n} &= x_{i.} - (n-1)\sqrt{\frac{1}{(n-1)n}}\, v_{i,n-1}
\end{aligned}
\tag{49}
$$

where $v_{i,1}, v_{i,2}, \cdots, v_{i,n-1}$ are new variables satisfying the identity

$$
\sum_{j=1}^{n-1} v_{i,j}^2 = \sum_{j=1}^{n} (x_{i,j} - x_{i.})^2. \tag{50}
$$

We transform them further by putting

$$
\begin{aligned}
v_{i,1} &= \sqrt{n}\, S_i \cos z_{i,n-2} \cos z_{i,n-3} \cdots \cos z_{i,2} \cos z_{i,1} \\
v_{i,2} &= \sqrt{n}\, S_i \cos z_{i,n-2} \cos z_{i,n-3} \cdots \cos z_{i,2} \sin z_{i,1} \\
v_{i,3} &= \sqrt{n}\, S_i \cos z_{i,n-2} \cos z_{i,n-3} \cdots \sin z_{i,2} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
v_{i,n-1} &= \sqrt{n}\, S_i \sin z_{i,n-2}
\end{aligned}
\tag{51}
$$

with the z's varying as follows

$$
\begin{aligned}
0 \le z_{i,1} &< 2\pi \\
-\pi/2 \le z_{i,j} &\le \pi/2
\end{aligned}
\qquad \text{for } j = 2, 3, \cdots, n-2. \tag{52}
$$

Of course, instead of the S_i we should put their expressions in terms of T_1 and the u's into (51). With the exception of a set of measure zero, which can be ignored, the formulae above determine a one-to-one transformation of the original space W of the x's into the space W' of $T_1, T_2, \cdots, T_{N+1}, u_1, \cdots, u_{N-1}$, and $z_{i,1}, z_{i,2}, \cdots, z_{i,n-2}$ for $i = 1, 2, \cdots, N$.

In calculating the joint probability law of all the new variables, we notice that, on the hypothesis tested, all the Nn original variables are mutually independent. Consequently, the transformations (49) and (51), which refer to separate groups of the $x_{i,j}$'s, corresponding to fixed values of i, could be carried

through separately. In doing so, we use formulae deduced elsewhere (see [5], pp. 38-39) directly and obtain

$$(53)\quad p(x_{i\cdot}, S_i, z_{i,1}, \cdots, z_{i,n-2}) = \left(\frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\right)^n S_i^{n-2} e^{-\frac{1}{2}n(S_i^2+(x_{i\cdot}-\xi_i)^2)/\sigma^2} \prod_{j=2}^{n-2} \cos^{j-1} z_{i,1}.$$

It follows that

$$\begin{aligned}(54)\quad & p(x_{1\cdot}, \cdots, x_{N\cdot}, S_1, \cdots, S_N, z_{i,1}, \cdots, z_{N,n-2}) \\ &= \prod_{i=1}^{N} p(x_{i\cdot}, S_i, z_{i,1}, \cdots, z_{i,n-2}) \\ &= \left(\frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\right)^{Nn} e^{-\frac{1}{2}n\sum_{i=1}^{N}(x_{i\cdot}-\xi_i)^2/\sigma^2} \prod_{i=1}^{N} S_i^{n-2} e^{-\frac{1}{2}nS_i^2/\sigma^2} \prod_{k=1}^{N}\prod_{j=2}^{n-2} \cos^{j-1} z_{k,j}.\end{aligned}$$

We now wish to introduce T_1 and the u_i instead of the S_i's. Since all other variables remain unchanged the Jacobian of this transformation reduces to that of (47). Simple calculations show that

$$(55)\quad \left|\frac{\partial(S_1, S_2, \cdots, S_N)}{\partial(T_1, u_1, \cdots, u_{N-1})}\right| = (\tfrac{1}{2})^N T_1^{\frac{1}{2}(N-2)} \left(1 - \sum_{i=1}^{N-1} u_i\right)^{-\frac{1}{2}} \prod_{i=1}^{N-1} u_i^{-\frac{1}{2}}.$$

Using this expression and substituting (47) in (54) we finally obtain

$$\begin{aligned}(56)\quad & p(x_{1\cdot}, \cdots, x_{N\cdot}, T_1, u_1, \cdots, u_{N-1}, z_{1,1}, \cdots, z_{N,n-2}) \\ &= (\tfrac{1}{2})^N \left(\frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\right)^{Nn} e^{-\frac{1}{2}n\sum_{i=1}^{N}(x_{i\cdot}-\xi_i)^2/\sigma^2} T_1^{\frac{1}{2}N(n-1)-1} \\ &\quad \cdot e^{-\frac{1}{2}nT_1/\sigma^2} \left(\left(1 - \sum_{i=1}^{N-1} u_i\right) \prod_{i=1}^{N-1} u_i\right)^{\frac{1}{2}(n-3)} \prod_{k=1}^{N}\prod_{j=2}^{n-2} \cos^{j-1} z_{k,j}.\end{aligned}$$

To obtain the relative probability law of $u_1, u_2, \cdots, u_{N-1}, z_{1,1}, \cdots, z_{N,n-2}$ given T_1 and the $T_{i+1} = x_{i\cdot}$, we must calculate $p(T_1, T_2, \cdots, T_{N+1})$ and divide expression (56) by it. Of course, $p(T_1, T_2, \cdots, T_{N+1})$ is obtained from (56) by integrating over the whole of $W'(\varphi)$, that is, for all other variables between the extreme limits of their variation. As these limits are independent of the values of $T_1, T_2, \cdots, T_{N+1}$, the result will be

$$(57)\quad p(T_1, T_2, \cdots, T_{n+1}) = ce^{-\frac{1}{2}n\sum_{i=1}^{N}(x_{i\cdot}-\xi_i)^2/\sigma^2} T_1^{\frac{1}{2}N(n-1)-1} e^{-\frac{1}{2}nT_1/\sigma^2}$$

where c denotes a constant. Thus

$$\begin{aligned}(58)\quad & p(u_1, \cdots, u_{N-1}, z_{1,1}, \cdots, z_{N,n-2} \mid T_1, \cdots, T_{N+1}) \\ &= c_1 \left(\left(1 - \sum_{i=1}^{N-1} u_i\right) \prod_{i=1}^{N-1} u_i\right)^{\frac{1}{2}(n-3)} \prod_{k=1}^{N}\prod_{j=2}^{n-2} \cos^{j-1} z_{k,j}\end{aligned}$$

with the region of variation $W'(\varphi)$ limited by the following inequalities

$$0 \leq u_i, \qquad \sum_{i=1}^{N+1} u_i \leq 1$$

$$0 \leq z_{k,1} < 2\pi \qquad \text{for } k = 1, 2, \cdots, N, \tag{59}$$

$$-\pi/2 \leq z_{k,j} \leq \pi/2 \qquad j = 2, 3, \cdots, n-2.$$

Since (58) integrated over $W'(\varphi)$ is identically unity, c_1 is a purely numerical constant.

Now to construct any region w similar to the sample space with respect to σ, $\xi_1, \cdots, \xi_N$, we must select, separately for each and all systems (φ) of values of $T_1, T_2, \cdots, T_{N+1}$, a region $w'(\varphi)$, part of $W'(\varphi)$ as defined by (59), with the sole restriction that

$$\int \cdots \int_{w'(\varphi)} p(u_1, \cdots, u_{N-1}, z_{1,1}, \cdots, z_{N,n-2} \mid T_1, \cdots, T_{N+1}) \cdot du_1, \cdots, dz_{N,n-2} = \epsilon. \tag{60}$$

Obviously, there is an infinity of ways of selecting any single one of such regions. For example, we could let the u's vary as indicated in (59) and limit the z's by

$$0 \leq z_{1,1} \leq a, \quad -a \leq z_{k,j} \leq a \quad (k = 1, 2, \cdots, N; j = 2, 3, \cdots, n-2) \tag{61}$$

where a is chosen so that (60) is satisfied. This choice of $w'(\varphi)$ may correspond to one particular system of values of $T_1, T_2, \cdots, T_{N+1}$ and no other. Again, the same region (61) may be chosen to serve for all systems of values of the T's. In this case, the region $w = \sum_{\varphi} w(\varphi)$ might be described as cylindrical. Any such region w will control errors of the first kind in testing H to the same level of significance ϵ and, as far as these errors alone are concerned, each of these regions is of equal value. Whatever the choice of regions $w'(\varphi)$ or $w(\varphi)$, the test of H will consist of (1) observing the values of the $x_{i,j}$'s, (2) calculating the corresponding value of $T_1, T_2, \cdots, T_{N+1}$, the u's, and the z's, and (3) noting whether the point with coordinates $u_1, u_2, \cdots, u_{N-1}, z_{1,1}, \cdots, z_{N,n-2}$ falls in the region $w(\varphi)$ chosen to correspond to the observed values of $T_1, T_2, \cdots, T_{N+1}$. Of course, in practical cases, the choice of $w'(\varphi)$ for one system of values of the T's will not be quite unconnected with that for others. On the contrary, there will probably be some more or less simple rule connecting $w'(\varphi)$ with the corresponding systems of the T's. As a result, the actual machinery of the test will be much simpler than that described above and will consist of the calculation of only a very few functions of the x's and in checking some simple inequalities.

Now our purpose is to select a region from the infinite family $F(\epsilon)$ of all regions similar to the sample space with respect to $\sigma, \xi_1, \cdots, \xi_N$ which we judge most satisfactory for controlling errors of the second kind. Roughly speaking,

this region will have to be such that, if the hypothesis H is not true, the observed point E will fall in this particular region as frequently as possible, in general. Here we come to the necessity of specifying the ways in which we expect the hypothesis H to be untrue. It may be untrue in an infinite number of ways. For example, the values of the σ's may (1) be equally distributed over any given range, (2) may fall into just two groups $\sigma_i = 1$ and $\sigma_i = 2$, or (8) all σ_i's except the last may have the same value σ while the last is 10σ, and so forth. Any such assumption will be called an hypothesis alternative to H. It is obvious that the probability of E falling in any given region w will be different for each of them. Therefore, if we wish to deduce a test which will detect the falsehood of the hypothesis tested frequently, we must analyse the practical cases where the test is to be applied and guess the ways in which the hypothesis tested is usually wrong. Then we can deduce a test which will be, in one sense or another, most sensitive to the assumed deviations from the hypothesis tested. Needless to say, our guess may be right or wrong In the latter case, an increased volume of observational material may demonstrate its fallacy and suggest the necessary modifications. In any case, it is important to know exactly the class of alternatives for which our test is, in some particular way, the best.

5. **The set of hypotheses alternative to H.** Let us consider the routine analyses made at some laboratory and try to discover the circumstances likely to cause variation in their accuracy. First of all, we may think of assignable causes such as a change in personnel, apparatus, or accommodation. These and similar causes are likely to produce lasting effects; the test of the hypothesis that they did not reduces to one of the equality of only two σ's. An easy application of known theory [20] shows that the familiar F or z test is unbiased of type B_1, which means that it is preferable to any other. Consequently, situations of this kind and also similar one for which the L_1 test is applicable [9], need not be considered here, so that we may concentrate on cases where there is no directly assignable cause of variation in the accuracy of the analyses. Assume then that the personnel, the apparatus, the accommodation, etc., remain the same. Now the accuracy of analyses depends on a multitude of causes evading identification, such as changes in the efficiency of the workers. In principle, they try to have the highest, and therefore a constant, level of accuracy. Uncontrollable circumstances cause some fluctuations about a certain average and we expect that small deviations from this average will occur more frequently than large ones. With this in mind, the author feels that it would be appropriate to expect that variations in accuracy, if any, will have a random character so that any σ_i referring to one particular group of analyses, or any monotonic function of that σ_i could be considered as an essentially positive random variable, having some unimodal probability law. To make the problem of the best test sufficiently specific, we must specify this law entirely. Here we face a somewhat embarassing freedom of choice. For lack of more precise information as to the random variability of σ_i, we guide ourselves by considerations of ease in

calculations. From this point of view it is convenient to consider the variable

$$h = \sigma^{-2} \tag{62}$$

and assume that, within a given period of time which is not too long, when the conditions in a laboratory are sensibly constant, it is varying according to the law

$$p(h) = \beta^{\alpha} h^{\alpha-1} e^{-\beta h}/\Gamma(\alpha) \quad \text{for} \quad 0 < h, \tag{63}$$

where α and β are unknown non-negative constants. It is useful to express these constants in terms of two new ones which have an obvious interpretation. h_0, the expectation of h, and ν, the square of the coefficient of variation of h. Easy calculations give

$$\alpha = 1/\nu, \qquad \beta = 1/h_0\nu. \tag{64}$$

Now $p(h)$ has the form

$$p(h) = \frac{1}{(h_0\nu)^{1/\nu}\Gamma(1/\nu)} h^{(1/\nu)-1} e^{-h/h_0\nu}. \tag{65}$$

We note that when $\nu \to 0$ the probability law (65) tends to a limiting discontinuous form with $P\{h = h_0\} = 1$. This corresponds to the hypothesis H that we wish to test. The type of law represented by (65) is known to be rather flexible. Consequently, we may easily assume that even though the true variability of h (or σ) does not exactly correspond to (65), there will be a system of values of h_0 and ν for which the difference between the true law and (65) will not be large. Therefore, a test which is particularly sensitive to deviations of ν from zero with law (65) will be reasonably sensitive in real practical cases. However, this is an assumption by the author. But it is subject to test and this will be done below.

Formula (63) represents the hypothetical probability law of the variable h which is not directly observable. We must use this formula to obtain the probability law of the observable x's alternative to (6), which corresponds to the hypothesis H being true. Using $h = 1/\sigma^2$, we write the relative probability law of $x_{i,1}, x_{i,2}, \cdots, x_{i,n}$ given h

$$p(x_{i,1}, \cdots, x_{i,n} \mid h) = \left(\frac{h}{2\pi}\right)^{n/2} e^{-\frac{1}{2}h\sum_1^n (x_{i,j}-\xi_i)^2}. \tag{66}$$

Multiplying (66) by (65) we obtain the joint probability law of h and the $x_{i,j}$'s referring to one group of analyses

$$p(h, x_{i,1}, \cdots, x_{i,n}) = \frac{1}{(2\pi)^{n/2}(h_0\nu)^{1/\nu}\Gamma(1/\nu)} h^{n/2+(1/\nu)-1} e^{-h\left(1/h_0\nu+\frac{1}{2}\sum_1^n (x_{i,j}-\xi_i)^2\right)}. \tag{67}$$

Integrating (67) with respect to h from zero to infinity, we obtain the absolute probability law of $x_{i,1}, x_{i,2}, \cdots, x_{i,n}$, all referring to the ith group of analyses. Assuming that the value of h in one group of analyses is independent of that in another, we obtain the joint probability law of all the Nn observable $x_{i,j}$'s by

simply multiplying the probability laws referring to particular groups of n of them. The result will depend on $N + 2$ unknown parameters, ξ_1, ξ_2, $\cdots$, ξ_N, h_0, and ν. As the last two will play a more important role than the others we shall denote the probability law by $p(E \mid h_0, \nu)$. Easy calculations give

$$(68) \qquad p(E \mid h_0, \nu) = \left(\frac{\Gamma(n/2 + 1/\nu)}{(2\pi)^{n/2}\Gamma(1/\nu)}\right)^N \frac{(h_0\nu)^{\frac{1}{2}Nn}}{\prod_{i=1}^{N}\left(1 + \frac{h_0\nu}{2}\sum_{j=1}^{n}(x_{i,j} - \xi_i)^2\right)^{n/2+1/\nu}}.$$

We easily check that for $\nu \to 0$ (68) approaches the law (6) with $h_0 = \sigma_*^{-2}$. Therefore, the problem that we shall treat below will be to assume that the observable x's follow (68) with *some* $h_0 > 0$ and *some* $\nu \geq 0$ and to test the hypothesis H that $\nu = 0$. More specifically, we shall try to choose among all the regions of the family $F(\epsilon)$, found in the preceding section the one over which the integral of the function (68) is, in general, the largest

Before doing so, it may be useful to exhibit some experimental evidence in favor of the assumption that, if σ is not constant in some conditions of analysis or measurement, then it varies in such a way that the variability of the x's has at least some characteristics appropriate to (68).

Introduce the notation

$$(69) \qquad \omega_i = nS_i^2 = \sum_{j=1}^{n}(x_{i,j} - x_{i\cdot})^2.$$

Using transformations (49), (50), and (69), successively, we easily deduce the probability law of ω_i

$$(70) \qquad p(\omega_i) = \frac{(h_0\nu/2)^{\frac{1}{2}(n-1)}\Gamma(\frac{1}{2}(n-1) + 1/\nu)}{\Gamma(\frac{1}{2}(n-1))\Gamma(1/\nu)} \frac{\omega_i^{\frac{1}{2}(n-3)}}{(1 + \frac{1}{2}h_0\nu\omega_i)^{\frac{1}{2}(n-1)+1/\nu}}.$$

If the hypothesis we have made about the variability of h, as expressed by (65), is true in any particular case then the sums of squares (69), referring to each particular group of analyses, are distributed according to (70). The reverse is not necessarily true, of course, but it is comforting that a check of the above in a number of broadly divergent circumstances gives satisfactory results. By applying the transformation $1 + h_0\nu\omega_i/2 = t^{-1}$, the integral of (70) is easily reduced to an Incomplete Beta function whence Pearson's tables [24] provide an easy means of calculating the theoretical probability that ω_i is within any given limits.

Table I gives several observed distributions of the sums ω together with their expected ones, calculated from (70) with the values of h_0 and ν fitted by the method of moments. The last lines give particulars of the application of the χ^2 test for goodness of fit.

The origin of the data used to compile Table I is as follows:

For the data providing frequency distributions numbered 1 and 2, the author is deeply indebted to Professor Raymond T. Birge. The methods of measurement and their purpose are explained in the publications [25] and [26], respec-

TABLE I

Comparison of empirical distributions of ω with those calculated from [illegible]

Number	1		2		3		4		5	
Author or Source of Data	R. T. Birge		R. T. Birge		K. Rusgeyn-ski and Sons, Ltd		A. A. Michel-son, F. G. Pease, and F. Pearson		W. S. Stevens	
Kind of Measurement or Analysis	Strong Lines in the Band Spectra of Nitrogen		A Solar Spectrum Line		Sugar Content of Beets		Velocity of Light		Octane Rating	
	Frequency		Frequency		Frequency		Frequency		Frequency	
ω	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.
0–1	29·38	29	15·10	17	15·56	16	3·50	2	14·90	17
1–2	19·30	20	13·14	11	12·67	17	7·73	10	18·88	16
2–3	13·11	17	11·39	15	10·70	13	9·37	13	16 83	11
3–4	9·16	7	9·84	5	8·98	2	9·66	8	13·93	12
4–5	6·56	6	8·46	9	7·53	11	9·28	17	11·20	10
5–6	4·80	1	7·24	9	6·34	4	8·60	7	8·91	7
6–7	3·59	4	6·17	11	5·36	3	7·80	7	7·04	10
7–8	4 80	1	5·23	4	4·54	7	6·99	7	5·58	9
8–9		3	4·40	2	3·86	4	6·22	1	4·43	7
9–10		2	3·69	2	6·09	0	5·52	4	3·52	7
10–11	3·94	0	5·63	2		5	4·88	3	5·08	3
11–12		0		1	4·45	0	4·32	5		1
12–13		4	3·76	3		0	3·82	3		0
13–14	5·36	1		1		5	6·37	2	4·51	1
14–15		0		1	4·61	1		5		0
15–16		0	5·95	3		0	5·03	1		1
16–17		1		1		0		0	6·18	1
17–18		0		1		3	4·00	3		1
18–19		0		1	4·37	1		2		2
19–20		1		1		0		2		0
20–21		1				1	4·55	0		0
21–22		0				0		1		0
22–23		0			4·94	1		2		0
23–24		0				0	4·23	1		0
24–25		0				0		1		0
25–26		0				1		1		0
26–32		2				4	3·94	3		1
32–43						1	3·58	3		1
>43							3·61	6		
Total	100·00	100	100·00	100	100·00	100	123·00	123	120·99	121
χ^2	9·63		12·67		18·75		18·09		13·35	
Degrees of Freedom	7		10		11		18		10	
$P(\chi^2)$	·21		·24		·066		·45		·21	

The symbols } are used to indicate the groupings used in the calculation of the χ^2. The groupings were made so as to have the expected frequency in a class at least equal to 3.5.

tively. These papers also contain various compilations of the results of the measurements. However, the original single measurements, necessary for the present paper, are naturally unpublished and Professor Birge was kind enough to find them for the author in his records.

Frequency distribution No. 3 was compiled from a book of records of sugar beet trials carried out by Messrs K. Buszczyński and Sons, Ltd. in Górka Narodowa, Poland.

The 4th distribution was constructed from the original measurements of the velocity of light as published [27] by Michelson, Pease, and Pearson The measurements made during single days were treated as forming separate groups

Distribution No. 5 originated from repeated measurements of Octane Rating conducted by a refining company in California They were made accessible by Mr. Walter S. Svenson and it is a pleasure to express the author's deep gratitude to him

The number of observations in each column is not very large. It may be expected that if it were increased, the differences between the hypothetical distributions and the observed ones would become more apparent. It seems safe, however, to assume that in a number of instances the hypothesis as to the character of the variability of ω_i is not in very bad disagreement with the actual facts. It would be most interesting to have some more data on the subject

6. **The best critical region for testing H against a particular alternative.** It seems unquestionable that the most desirable test of any hypothesis is the uniformly most powerful test (U M P. Test) with respect to the whole class of simple hypotheses alternative to the one which is being tested. Denote by H the hypothesis tested, by h any simple admissible hypothesis alternative to H, and by Ω the set of all h's. If w_0 is the critical region corresponding to the U. M. P. Test, then w_0 has these properties:

(71) (1) $$P\{E \epsilon w_0 \mid H\} \equiv \epsilon.$$

(2) If w is any other region such that $P\{E \epsilon w \mid H\} \equiv \epsilon$ then

(72) $$P\{E \epsilon w_0 \mid h\} \geq P\{E \epsilon w \mid h\},$$

whatever be $h \epsilon \Omega$.

Following the known method [18], we shall see whether a test of the hypothesis H considered in the preceding sections exists which is a U. M. P. Test with respect to the whole class of admissible hypotheses that specify the probability laws (68) with any $h_0 > 0$ and $\nu > 0$.

The method consists of considering one particular alternative hypothesis h', that is, one particular set of values of $h_0 > 0$ and $\nu > 0$ and finding the best critical region $w_{h_0,\nu}$ for testing H against h'. If this region appears to depend on ν and/or on h_0 then there is no U. M. P. Test The region $w_{h_0,\nu}$ is found by determining, for each system (φ) of $T_1, T_2, \cdots, T_{N+1}$ separately, a part $w_{h_0,\nu}(\varphi)$ determined by the inequality

(73) $$p(E \mid h_0, \nu) \geq k(\varphi)p(E \mid H)$$

where $k(\varphi)$ is a function of T_1, T_2, $\cdots$, T_{N+1} so determined that the relation (60) is satisfied. Substituting (66) and (68) in (73), taking the logarithm of both sides, and combining all terms which are constant or depend only on T_1, T_2, $\cdots$, T_{N+1}, we have

$$(74) \quad \sum_{i=1}^{N} \log (1 + \tfrac{1}{2}h_0\nu n(S_i^2 + (T_{i+1} - \xi_i)^2)) \leq k_1(T_1, \cdots, T_{N+1}). \quad (\text{say } v).$$

Clearly, for T_1, T_2, $\cdots$, T_{N+1} fixed, this inequality imposes a restriction on the variability of u_1, u_2, $\cdots$, u_{N+1} while $z_{1,1}$, $\cdots$, $z_{N,n-2}$ are allowed to vary indiscriminately within the extreme limits (52). But the region $w_{k_1}(v)$ determined by (74) also depends on the product $h_0\nu$. Therefore, there is no uniformly most powerful test for testing H against any and all simple alternatives specifying (68).

7. **A critical region of an unbiased type.** There seems to be no grounds for dissention that when a U. M. P. Test exists and is readily applicable, it is preferable to any other test, but the situation is quite different when there is no U. M. P. Test. In such cases, practical considerations may suggest a variety of requirements for a second best test of the hypothesis. Among these, we may suggest the following considerations:

Fix, for a moment, the values of h_0, ξ_1, $\cdots$, ξ_N, take any region w of the family $F(\epsilon)$, and consider the probability of E falling in w as a function of ν only. This is called the power function

$$(75) \qquad \beta(\nu \mid w) = \int \cdots \int_w p(E \mid h_0, \nu)\, dx_{1,1} \cdots dx_{N,n}$$

Here, of course, $\nu \geq 0$. Because of the properties of regions belonging to $F(\epsilon)$ we have $\beta(0 \mid w) \equiv \epsilon$. If $\nu > 0$, the value of $\beta(\nu \mid w)$ represents the corresponding probability of the test (based on w) discovering the falsehood of H. It is obviously desirable to have this probability as large as possible. In any case, it should be greater than ϵ. This last restriction is known as that of unbiasedness [19], [20], [28]. Further, since it is impossible to maximize $\beta(\nu \mid w)$ for *all values* of ν, we must choose those for which it is most desirable, in our opinion, to concentrate our efforts to increase $\beta(\nu \mid w)$. One possible point of view is that these values should be very close to the hypothetical value $\nu = 0$. For if ν is considerably larger than zero, we may argue that there will be no need to apply any refined statistical test to detect the falsehood of H. Of course, this argument has no mathematical character and its general acceptance is not suggested. In fact, we may argue that if ν is greater than zero but very small, it will be almost impossible to detect the falsehood of H by any test and, therefore, our efforts should be concentrated on values of ν which are of considerable size.

These are considerations of non-mathematical character; the role of mathematical statistics is limited to devising tests and elucidating their properties. If these last are understood by practical statisticians, each may choose according

to his problem. Note that what could be termed the "properties" of a test are summarized in the power function $\beta(\nu \mid w)$ with its relation to the power functions of other possible tests of the same hypothesis.

In this paper we shall deal with tests particularly sensitive to small deviations of ν from its hypothetical value $\nu = 0$. In this respect, our first trial is to find a region w_0, belonging to the family $F(\epsilon)$ and satisfying the condition

$$\frac{\partial\beta(\nu \mid w_0)}{\partial\nu}\Big]_{\nu=0} \geq \frac{\partial\beta(\nu \mid w)}{\partial\nu}\Big]_{\nu=0}, \tag{76}$$

where w is any other region belonging to the same family $F(\epsilon)$.

Because of the peculiar structure of the regions belonging to $F(\epsilon)$, the problem is immediately reduced to finding regions $w_0(\varphi)$. According to theory explained elsewhere [18] these should satisfy the condition

$$\frac{\partial p(E \mid h_0, \nu)}{\partial\nu}\Big]_{\nu=0} \geq k(T)p(E \mid H), \tag{77}$$

where $k(T)$ depends on $T_1, T_2, \cdots, T_{N+1}$ only and is determined to satisfy the condition of similarity (60). Condition (77) is equivalent to

$$\frac{\partial \log p(E \mid h_0, \nu)}{\partial\nu}\Big]_{\nu=0} \geq k(T). \tag{78}$$

Taking the logarithm of (68), differentiating with respect to ν, putting ν equal to zero, substituting in (78), and combining all the terms which are constant on $W(\varphi)$ into a single term which we may write as $\frac{1}{8}h_0^2k_1(T)$, we have

$$\sum_{i=1}^{N} (S_i^2 + (T_{i+1} - \xi_i)^2)^2 \geq k_1(T). \tag{79}$$

We note that condition (79) determining, so to speak, the shape of the region $w_0(\varphi)$ does not imply any restriction on the variability of the z's but only on the u's. However, the region $w_0(\varphi)$ as determined by (79) has the disadvantage of being dependent on the values of the ξ_i. Since these are not specified by the hypothesis tested, we are not able to determine the critical regions belonging to the family $F(\epsilon)$ and maximizing the derivative $\partial\beta(\nu \mid w)/\partial\nu]_{\nu=0}$. The region which does so for some particular system $\xi_1', \xi_2', \cdots, \xi_N'$ of values of the ξ's will lose this property if the system of values of the ξ's is appropriately changed. Therefore, our choice of the region maximizing the derivative of the power function at $\nu = 0$ should be made not from the whole family $F(\epsilon)$ but from a subfamily $F_1(\epsilon)$ composed only of such regions which also possess the supplementary property that

$$\frac{\partial\beta(\nu \mid w)}{\partial\nu}\Big]_{\nu=0} = \text{constant} \tag{80}$$

has a value independent of $\xi_1, \xi_2, \cdots, \xi_N$. The determination of this subfamily $F_1(\epsilon)$ embracing all such regions is an interesting problem. Until it is

solved, we use an obvious subfamily $F_2(\epsilon)$ of regions w which have the desired property, but we do not know whether or not $F_2(\epsilon)$ contains all such regions.[2]

The family $F_2(\epsilon)$ is defined as consisting of those regions belonging to $F(\epsilon)$ which could be described as cylindrical with their generators parallel to the intersection of $T_{i+1} = x_{i+1} = \text{constant}$, for $i = 1, 2, \cdots, N$. In other words and more precisely, a region w of the family $F_2(\epsilon)$ belongs to $F(\epsilon)$ only if the question of its including a given point E depends on $Nn - N$ of the coordinates, namely on $T_1, u_1, \cdots, u_{N-1}, z_{1,1}, \cdots, z_{N,n-2}$ and not on $T_2, T_3, \cdots, T_{N+1}$.

We easily show that any region w belonging to $F_2(\epsilon)$ possesses the property that its power function is independent of the ξ_i's. Denote by w' the set of systems of values of $T_1, u_1, \cdots, u_{N-1}, z_{1,1}, \cdots, z_{N,n-2}$ corresponding to points included in any given region w of the family $F_2(\epsilon)$. We see that the power function $\beta(\nu \mid w)$, equal to the integral of (68) over w, can be calculated by using the transformations (47), (49), and (51). Then the region of integration for $T_1, u_1, \cdots, u_{N-1}, z_{1,1}, \cdots, z_{N,n-2}$ is what we have just denoted by w' and the integrations for $T_{i+1} = x_{i+1}$ extend from $-\infty$ to $+\infty$ irrespective of the fixed values of the other variables. These integrations are easily carried out by substituting

$$\tfrac{1}{2}nh_0\nu(x_{i+1} - \xi_{i+1})^2 = (1 + \tfrac{1}{2}nh_0\nu S_i^2)t_i^2. \tag{81}$$

The final result is

$$\beta(\nu \mid w) = \int \cdots \int_{w'} p(T_1, u_1, \cdots u_{N-1}, z_{1,1}, \cdots z_{N,n-2})\, dT_1 \cdots dz_{N,n-2} \tag{82}$$

Here

$$p(T_1, u_1, \cdots u_{N-1}, z_{1,1}, \cdots z_{N,n-2}) = c(\nu)\Phi(T_1, u, z) \Big/ \prod_{i=1}^{N} (1 + \tfrac{1}{2}nh_0\nu S_i^2)^{\frac{n-1}{2}}, \tag{83}$$

where $c(\nu)$ denotes a constant depending on ν, $\Phi(T_1, u, z)$ denotes a function of all the $N(n-1)$ variables involved, independent of ν, and S_i^2 denotes expressions (47) for short. We see that (82) is independent of the ξ_i's.

Since the region w belongs to $F(\epsilon)$, it is composed of sections $w(\omega)$ selected separately on each hypersurface $T_1 = \text{constant}$ and $T_{i+1} = \text{constant}$, $i = 1, 2, \cdots, N$. Because of the definition of the family $F_2(\epsilon)$, the sections $w(\omega)$ are independent of $T_2, T_3, \cdots, T_{N+1}$ so that each of them can be selected only in accordance with the value of T_1. Therefore, we may denote them by $w(T_1)$. As far as property (80) is concerned, the choice is arbitrary. But the property of similarity requires the fulfillment of condition (60) which, in the present case, reduces to

$$\int \cdots \int_{w(T_1)} p(u_1, \cdots u_{N-1}, z_{1,1}, \cdots z_{N,n-2} \mid T_1, \cdots T_{N+1})\, du_1 \cdots dz_{N,n-2} = \epsilon \tag{84}$$

[2] Regions with the property (80) and belonging to $F(\epsilon)$ but not to $F_2(\epsilon)$ exist. Probably however, each of them differs from one of the regions of $F_2(\epsilon)$ by a set of measure zero only.

Applying the method already used, we find that sections $\bar{w}(T_1)$ of the region $\bar{w}$ belonging to $F_2(\epsilon)$ and maximizing the derivative $\partial\beta(\nu \mid w)/\partial\nu]_{\nu=0}$ are determined, separately for each value of T_1, by the inequality

$$\frac{\partial \log p(T_1, u_1, \cdots u_{N-1}, z_{1,1}, \cdots z_{N,n-1})}{\partial \nu}\Bigg]_{\nu=0} \geq k_2(T_1) \tag{85}$$

where $k_2(T_1)$ denotes a function of T_1 determined to satisfy (84).

Substituting (83) in (85) we easily find that this condition is equivalent to

$$\zeta \equiv \sum_{i=1}^{N-1} u_i^2 + \left(1 - \sum_{i=1}^{N-1} u_i\right)^2 \geq k_3(T_1) \tag{86}$$

where, again, $k_3(T_1)$ is determined for each particular value of T_1 to satisfy (84). As (86) does not imply any restrictions on the variability of $z_{1,1}, z_{1,2}, \cdots, z_{N,n-2}$, the integrations for the z's while calculating (84) must be carried out over the extreme limits (52). This will reduce the integrand to the relative probability law of $u_1, u_2, \cdots, u_{N-1}$ given all the T's. This law is easily calculated from (58) and is

$$\begin{aligned} &p(u_1, u_2, \cdots u_{N-1} \mid T_1, T_2, \cdots T_{N+1}) \\ &\qquad = \frac{\Gamma(\frac{1}{2}N(n-1))}{\Gamma^N(\frac{1}{2}(n-1))}\left(\left(1 - \sum_{i=1}^{N-1} u_i\right)\prod_{i=1}^{N-1} u_i\right)^{\frac{1}{2}(n-3)} \\ &\qquad = p(u_1, u_2, \cdots u_{N-1}) \end{aligned} \tag{87}$$

As (87) is independent of $T_1, T_2, \cdots, T_{N+1}$, it is also the absolute probability law of the u's and hence $k_3(T_1)$ is independent of T_1. In accordance with the notation adopted for the left side of (86), namely ζ, and since the choice of $k_3(T_1)$ depends on ϵ, n, and N, we may use ζ_ϵ instead of $k_3(T_1)$. Then the region $\bar{w}$ is determined by the inequality

$$\zeta = \sum_{i=1}^{N-1} u_i^2 + \left(1 - \sum_{i=1}^{N-1} u_i\right)^2 \geq \zeta_\epsilon \tag{88}$$

or, returning to the original variables, by the inequality

$$\zeta \equiv \sum_{i=1}^{N} S_i^4 \Big/ \left(\sum_{i=1}^{N} S_i^2\right)^2 \geq \zeta_\epsilon \tag{89}$$

where ζ_ϵ is the root of the equation

$$\frac{\Gamma(\frac{1}{2}N(n-1))}{\Gamma^N(\frac{1}{2}(n-1))} \int \cdots \int_{\zeta \geq \zeta_\epsilon} \left(\left(1 - \sum_{i=1}^{N-1} u_i\right)\prod_{i=1}^{N-1} u_i\right)^{\frac{1}{2}(n-3)} du_1 \cdots du_{N-1} = \epsilon \tag{90}$$

This region $\bar{w}$ has the following property: of all the regions belonging to the family $F_2(\epsilon)$, the derivative of the power function of $\bar{w}$ at the point $\nu = 0$ is the greatest. Thus, as far as the values of ν close to zero are concerned, we may say that, for testing H, $\bar{w}$ is the most powerful critical region in the family $F_2(\epsilon)$.

8. **Methods of determining ζ_ϵ.** To calculate ζ_ϵ accurately we must calculate the integral probability law of ζ, that is to say,

$$P\{\zeta < z\} = \int \cdots \int_{\zeta<z} p(u_1, \cdots u_{N-1})\, du_1 \cdots du_{N-1} \tag{91}$$

for any z. The author was not able to achieve this. Therefore some methods of approximation had to be looked for. This task becomes somewhat simplified by noting that in most practical problems N will be very large, in the hundreds or thousands, while n will probably not exceed 5.

To start, we notice that the range of ζ is limited by

$$1/N \leq \zeta \leq 1. \tag{92}$$

The easiest way to see this is to look for maxima and minima of the sum

$$X = \sum_{i=1}^{N} S_i^4 \tag{93}$$

subject to the restriction that

$$\sum_{i=1}^{N} S_i^2 = T_1 \tag{94}$$

We then easily find that

$$T_1^2/N \leq X \leq T_1^2 \tag{95}$$

and (92) follows directly.

Since ζ is a polynomial of the second order in the u's, we may consider its moments. These will be functions of the expectations of the products $\prod_{i=1}^{N} u_i^{k_i}$, where, for short, $u_N = 1 - \sum_{i=1}^{N-1} u_i$. Using (87) we easily find that

$$E\left(\prod_{i=1}^{N} u_i^{k_i}\right) = \frac{\Gamma(\frac{1}{2}N(n-1))}{\Gamma\left(\frac{1}{2}N(n-1) + \sum_{i=1}^{N} k_i\right)} \prod_{i=1}^{N} \frac{\Gamma(\frac{1}{2}(n-1) + k_i)}{\Gamma(\frac{1}{2}(n-1))}. \tag{96}$$

In particular, if we let $(n-1)/2 = a$

$$E(u_i^2) = \frac{a(a+1)}{Na(Na+1)} \tag{97}$$

$$E(u_i^4) = \frac{a(a+1)(a+2)(a+3)}{Na(Na+1)(Na+2)(Na+3)} \tag{98}$$

$$E(u_i^2 u_j^2) = \frac{a^2(a+1)^2}{Na(Na+1)(Na+2)(Na+3)}. \tag{99}$$

Consequently and because $\zeta = \sum_{i=1}^{N} u_i^2$, we have

$$E(\zeta) = \mu_1' = (a+1)/(Na+1) \tag{100}$$

$$E(\zeta^2) = \mu_2' = \sum_{i=1}^{N} E(u_i^4) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} E(u_i^2 u_j^2)$$
$$= \frac{(a+1)(a+2)(a+3)}{(Na+1)(Na+2)(Na+3)} + \frac{(N-1)a(a+1)^2}{(Na+1)(Na+2)(Na+3)}. \tag{101}$$

The variance σ_ζ^2 of ζ is therefore

$$\sigma_\zeta^2 = \frac{2a(a+1)(N-1)}{(Na+1)^2(Na+2)(Na+3)}. \tag{102}$$

By a similar procedure we find that

$$E(\zeta^3) = \mu_3' = \frac{\begin{array}{r}(a+1)(a+2)(a+3)(a+4)(a+5)\\ +3(N-1)a(a+1)^2(a+2)(a+3)\\ +(N-1)(N-2)a^2(a+1)^3\end{array}}{(Na+1)(Na+2)(Na+3)(Na+4)(Na+5)} \tag{103}$$

$$E(\zeta^4) = \mu_4' = \frac{\begin{array}{r}\prod_{j=1}^{7}(a+j) + 4(N-1)a(a+1)\prod_{j=1}^{5}(a+j)\\ +3(N-1)a\prod_{j=1}^{3}(a+j)^2\\ +6(N-1)(N-2)a^2(a+1)^3(a+2)(a+3)\\ +(N-1)(N-2)(N-3)a^3(a+1)^4\end{array}}{\prod_{j=1}^{7}(Na+j)} \tag{104}$$

One possible method of approximating ζ, is to use the formulae above, together with the higher moments whose formulae are easy to deduce. Some convenient known distribution, say $p_0(\zeta)$, could be fitted to have its first two or three moments coincide with those of the unknown true distribution of ζ. We would then look for better approximations by means of the functions

$$p_m(\zeta) = p_0(\zeta)\sum_{j=1}^{m} A_j \pi_j \tag{105}$$

where the π_j's denote polynomials which are orthogonal and normal with respect to $p_0(\zeta)$ so that

$$\int \pi_j \pi_k p_0(\zeta)\, d\zeta = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k. \end{cases} \tag{106}$$

The constant coefficients A_j are formed to minimize the integral

$$\int \left(p(\zeta) - p_0(\zeta)\sum_{j=1}^{m} A_j \pi_j\right)^2 p_0^{-1}(\zeta)\, d\zeta. \tag{107}$$

They are expressible in terms of the known moments of $p(\zeta)$.

This is one possible way to approximate $p(\zeta)$ which would eventually lead to the computation of ζ_ϵ even for small values of N.

Remembering that we are concerned with large N's, we can prove that the normalized distribution of ζ, that is, the distribution of

$$\frac{\zeta - \mu_1'}{\sigma_\zeta} \tag{108}$$

tends to be normal as $N \to \infty$. However, the process of tending to the limit is rather slow as may be seen from the following table of K. Pearson's β_1 and β_2.

TABLE II

Frequency constants of the distribution of ζ

n	N	μ_1'	σ_ζ	β_1	β_2
3	100	.0198	.001922	.8652	5.042
3	200	.0099	.000693	.4618	4.211
3	400	.0050	.000248	.2410	3.587

Because of this and also because the proof that the distribution of (108) tends to normality is not very straightforward, we shall not reproduce it. But it may be well to point out that the cause of this slowness in tending to the limit lies in the skewness of the distribution of each particular u_i and in the mutual dependency of all the u_i's.

The most promising method seems to be the following. First consider the two sums

$$T_1 = \sum_{i=1}^{N} S_i^2 \quad \text{and} \quad T_0 = \sum_{i=1}^{N} S_i^4. \tag{109}$$

Obviously, these two sums satisfy the conditions of the limiting theorem of S Bernstein [29], [30] and, therefore, as $N \to \infty$, their joint normalized distribution tends to a normal surface Also, we may expect the process of tending to the limit to be rapid in this case. If $p(T_0, T_1)$ denotes the limiting normal distribution, the probability that $\zeta \geq z$ can be approximately calculated by the integral

$$P\{\zeta \geq z\} = P\{T_0 \geq zT_1^2\} = \int_{-\infty}^{+\infty} dT_1 \int_{zT_1^2}^{\infty} p(T_0, T_1)\, dT_0. \tag{110}$$

To calculate the limiting distribution $p(T_0, T_1)$ we need only the expectations, say A and B, of T_1 and T_0 respectively, their standard errors, say σ_1 and σ_2, and their correlation coefficient R. These may be obtained from the moments of the S_i^2's.

Formula (110) can be used not only for tabulating the integral probability law of ζ and for determining ζ_ϵ, but also for an approximate calculation of the power function of the test. For, if the limiting probability law $p(T_0, T_1)$ is

calculated using the moments of S_i^2 calculated from (70) with some $\nu > 0$, then the integral (110) calculated with $z = \zeta_\epsilon$ gives us the probability $P\{\zeta \geq \zeta_\epsilon \mid \nu\}$ of the test detecting the falsehood of the hypothesis tested, that is, the power function.

To save space, we shall now calculate the constants A, B, σ_1, σ_2, and R as functions of $\nu \geq 0$. The values appropriate to the case when the hypothesis tested is true will then be obtained from the general formulae by the mere substitution of $\nu = 0$.

Since all the constants above depend on the expectations of S_i^{2k}, we use formula (70) to calculate them. Denoting the expectation of S_i^{2k} by μ_k, we have

$$\mu_k = \frac{2(nh_0\nu/2)^{\frac{1}{2}(n-1)}}{\mathrm{B}(1/\nu, \frac{1}{2}(n-1))} \int_0^\infty \frac{S^{2k+n-2}}{(1 + \frac{1}{2}nh_0\nu S^2)^{\frac{1}{2}(n-1)+1/\nu}}\, dS. \tag{111}$$

Introducing the new variable

$$1 + \tfrac{1}{2}nh_0\nu S^2 = t^{-1} \tag{112}$$

makes the integration straightforward and gives

$$\mu_k = \left(\frac{2}{nh_0\nu}\right)^k \frac{\Gamma((1/\nu) - k)\Gamma(\frac{1}{2}(n-1) + k)}{\Gamma(1/\nu)\Gamma(\frac{1}{2}(n-1))}. \tag{113}$$

This formula holds good if $1/\nu > k$. Otherwise the kth moment μ_k is divergent. So this approximate method of calculating the power function of the test is applicable only for $\nu < 25$.

Substituting $k = 1, 2, 3, 4$ in (113), we have

$$\begin{aligned}
\mu_1 &= \frac{1}{nh_0}\frac{n-1}{1-\nu} \\
\mu_2 &= \left(\frac{1}{nh_0}\right)^2 \frac{n^2-1}{(1-\nu)(1-2\nu)} \\
\mu_3 &= \left(\frac{1}{nh_0}\right)^3 \frac{(n^2-1)(n+3)}{(1-\nu)(1-2\nu)(1-3\nu)} \\
\mu_4 &= \left(\frac{1}{nh_0}\right)^4 \frac{(n^2-1)(n+3)(n+5)}{(1-\nu)(1-2\nu)(1-3\nu)(1-4\nu)},
\end{aligned} \tag{114}$$

and now we have

$$A = \frac{N}{nh_0}\frac{n-1}{1-\nu}, \qquad B = \frac{N}{(nh_0)^2}\frac{n^2-1}{(1-\nu)(1-2\nu)}, \tag{115}$$

$$\sigma_1^2 = \frac{N}{(nh_0)^2}\frac{(n-1)(2+\nu(n-3))}{(1-\nu)^2(1-2\nu)}, \tag{116}$$

$$\sigma_2^2 = \frac{2N}{(nh_0)^4}\frac{(n^2-1)(2+\nu(n-3))(2(n+2)-\nu(5n+7))}{(1-\nu)^2(1-2\nu)^2(1-3\nu)(1-4\nu)}, \tag{117}$$

$$R^2 = \frac{2(n+1)(1-2\nu)(1-4\nu)}{(2(n+2)-\nu(5n+7))(1-3\nu)}. \tag{118}$$

Inspecting formulae (115) to (118) makes us see that there is an advantage in substituting two new variables

$$t_1 = \frac{nh_0}{N(n-1)}\,T_1, \qquad t_2 = \frac{(nh_0)^2}{N(n^2-1)}\,T_0, \tag{119}$$

for T_1 and T_0. Their expectations, say ϑ_1 and ϑ_2, are

$$\vartheta_1 = \frac{1}{1-\nu}, \qquad \vartheta_2 = \frac{1}{(1-\nu)(1-2\nu)}. \tag{120}$$

Probably without any danger of confusion, the S.E.'s of t_1 and t_2 may be denoted by σ_1 and σ_2 also and we shall have

$$\begin{aligned} \sigma_1^2 &= \frac{2+\nu(n-3)}{N(n-1)(1-\nu)^2(1-2\nu)}, \\ \sigma_2^2 &= \frac{2(2+\nu(n-3))(2(n+2)-\nu(5n+7))}{N(n^2-1)(1-2\nu)^2(1-3\nu)(1-4\nu)}. \end{aligned} \tag{121}$$

Of course, the correlation coefficient of t_1 and t_2 is the same as that of T_1 and T_0, namely R. Obviously, the inequality $T_0 \geq zT_1^2$ is equivalent to $t_2 \geq z_1 t_1^2$ provided that

$$z = z_1 \frac{n+1}{N(n-1)}. \tag{122}$$

Now the problem of calculating (110) is reduced to finding

$$\begin{aligned} P\{\zeta \geq z\} &= P\{t_2 \geq z_1 t_1^2\} \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-R^2}} \iint_{t_2 \geq z_1 t_1^2} \exp\left[-\frac{1}{2(1-R^2)}\left\{\frac{(t_1-\vartheta_1)^2}{\sigma_1^2}\right.\right. \\ &\qquad \left.\left. - 2R\,\frac{(t_1-\vartheta_1)(t_2-\vartheta_2)}{\sigma_1\sigma_2} + \frac{(t_2-\vartheta_2)^2}{\sigma_2^2}\right\}\right] dt_1\,dt_2. \end{aligned} \tag{123}$$

We may conveniently see the workings of the test proposed by considering formula (123). First consider the case when the hypothesis tested is true. Both ϑ_1 and ϑ_2 reduce to unity. The region of highest frequency is around the point $t_1 = t_2 = 1$. If N is large then both σ_1 and σ_2 are small so that the region of significant frequency is rather small. The integral (123) is to be taken over the region above the parabola $t_2 = z_1 t_1^2$ passing through the origin of coordinates. When z_1 is small and the parabola passes far below the point $t_1 = t_2 = 1$, the probability $P\{\zeta \geq z\}$ will be close to unity. When $z_1 = 1$ this probability will be less than $\frac{1}{2}$ and it will diminish rapidly with further increases of z_1. Now suppose that we have found the value ζ_ϵ for which $P\{\zeta \geq \zeta_\epsilon \mid \nu = 0\} = \epsilon$ and consider what will happen to (123) when $z = \zeta_\epsilon$ if ν is increased. Clearly, neither of σ_1 and σ_2 nor R are very sensitive to slight changes in ν. Also ϑ_1 will not change very much. On the other hand, ϑ_2 will increase rather fast. The final

conclusion is that the whole frequency surface corresponding to the integrand in (123) will not change shape much but will shift to bring a greater amount of frequency into the region of integration.

To facilitate numerical calculations introduce

$$x = \frac{t_1 - \vartheta_1}{\sigma_1}, \qquad y = \frac{t_2 - \vartheta_2 - R\sigma_2(t_1 - \vartheta_1)/\sigma_1}{\sigma_2\sqrt{1 - R^2}}. \tag{124}$$

Now (123) may be rewritten as

$$P\{\zeta \geqq z\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}x^2} \frac{1}{\sqrt{2\pi}} \int_{y(x,z_1)}^{\infty} e^{-\frac{1}{2}y^2}\, dy \right\} dx \tag{125}$$

where

$$y(x, z_1) = \frac{z_1(\vartheta_1 + \sigma_1 x)^2 - \vartheta_2 - R\sigma_2 x}{\sigma_2\sqrt{1 - R^2}}. \tag{126}$$

Using formulae (125), (126) and (119) to (122), the following numerical values were obtained.

TABLE III

$n = 3, \quad N = 100, \quad \nu = 0.$

z_1	$P\{\zeta \geq z \mid \nu = 0\}$
.8	.9126
.9	.7305
1.0	.4905
1.1	.2847
1.2	.1495
1.3	.0730
1.4	.0335
1.5	.0148
1.6	.00644
1.7	.00288
1.34450	.05000
1.54563	.01000

TABLE IV

Power of the test for $n = 3$ *and* $N = 100$.

ϵ	ζ_ϵ	$\nu = .01$	$\nu = .16$
.05	.02689	.05823	.37482
.01	.03091	.01234	.10699

The figures above are only approximate and we realize that the greater the value of ν the less satisfactory is the approximation of the power function. A check of the goodness of the approximation and, if it proves satisfactory, a few

numerical tables for practical applications of the test must be postponed to another publication.

It is a pleasure to record the author's indebtedness to Miss Elizabeth Scott and also to Miss Julia Bowman for carrying out all the numerical work connected with the present paper.

REFERENCES

[1] J. Neyman, *Jour. Roy. Stat. Soc.*, Vol. 97 (1934), p. 558
[2] J. Neyman, *Phil. Trans. Roy. Soc London*, Vol. 236-A (1937), p. 333.
[3] "Student", *Biometrika*, Vol. 19, (1927), p. 151.
[4] J. Przyborowski, *Roczniki Nauk Rolniczych i Leśnych*, Vol 30, (1933), p. 1.
[5] J. Neyman; *Lectures and conferences on mathematical statistics*, Washington, D. C. (1937).
[6] J Berkson, *Jour. Am Stat. Assn*, Vol. 33, (1938), p. 526.
[7] J Berkson, *Jour. Am. Stat. Assn*, Vol. 35, (1940), p. 362.
[8] J. Neyman, *Annals of Math. Stat*, Vol. 11 (1940), p. 478.
[9] J. Neyman and E. S. Pearson, *Bull. Int. Acad. Polon. Sci. Cracovie*, (A) Vol. 6 (1931), p. 460.
[10] B. L. Welch, *Biometrika*, Vol. 27 (1935), p. 145.
[11] B. L. Welch, *Stat. Res. Memoirs*, Vol. 1 (1936), p. 52.
[12] U. S. Nair, *Stat. Res. Memoirs*, Vol. 1 (1936), p. 38.
[13] M. S. Bartlett, *Proc. Roy Soc*, Vol. 160-A (1937), p. 268.
[14] S. S Wilks and C. M. Thompson, *Biometrika*, Vol. 29 (1937), p. 124.
[15] D. T. Bishop and U S. Nair, *Jour. Roy. Stat. Soc. Suppl.*, Vol. 6 (1939), p. 89.
[16] E. J. G. Pitman, *Biometrika*, Vol. 31, (1939), p. 200.
[17] H. O Hartley, *Biometrika*, 31 (1940), p. 249.
[18] J Neyman and E. S. Pearson, *Phil Trans Roy Soc. London*, 231-A (1933), p. 289.
[19] J. Neyman and E. S Pearson, *Stat. Res. Memoirs*, Vol. 1 (1936), p. 1.
[20] J. Neyman, *Bull. Soc Math. France*, Vol. 63 (1935), p. 246.
[21] H. Hotelling, *Ann. Math. Stat.*, Vol. 11 (1940), p. 271.
[22] W. Feller, *Stat Res. Memoirs*, Vol. 2 (1938), p. 117.
[23] H Cramér and H. Wold, *Jour. London Math. Soc*, Vol. 11 (1936), p. 291.
[24] K. Pearson, *Tables of the Incomplete Beta Function*, Biometrika Office, University College, London, 1934
[25] R. T. Birge, *Astrophysical Journal*, Vol. 39 (1914), p. 50.
[26] R. T Birge, *Physical Review*, Vol. 40 (1932), p. 207.
[27] A. A. Michelson, F. G. Pease, and F. Pearson, *Astrophysical Journal*, Vol. 82 (1935), p 26
[28] J. Neyman and E. S Pearson, *Stat Res Memoirs*, Vol. 2 (1938), p. 25.
[29] S. Bernstein, *Math. Ann.*, Vol. 97 (1926), p. 44.
[30] W Kozakiewicz, *Ann. Soc. Polon. Math.*, Vol. 13 (1934), p. 24.

A CONCISE ANALYSIS OF CERTAIN ALGEBRAIC FORMS

By Franklin E. Satterthwaite

State University of Iowa, Iowa City, Iowa

Many of the statistics in common use are functions of homogeneous algebraic forms in the items of the sample. Among such statistics are the mean, a linear form; the variance, a quadratic form; and the product moment, a bilinear form. With the extension of the science, the mathematical statistician is faced with the study of more complex statistics and the associated algebraic forms and matrices The purpose of this paper is to set forth concise and efficient notations and methods which may be used in such analysis.

We shall borrow the essential features of our notation from differential geometry and tensor analysis. The Kroneker delta is defined as,

$$\delta_i^j = 1, \qquad i = j,$$
$$= 0, \qquad i \neq j.$$

The summation convention provides that summation be performed with respect to any index appearing twice in the same term. Thus,

$$x_i y^i = x_1 y^1 + x_2 y^2 + \cdots .$$

To extend the use of the summation convention, we shall frequently place indices on the numeral, 1. Thus,

$$1^i x_i = 1^1 x_1 + 1^2 x_2 + \cdots = x_1 + x_2 + \cdots .$$

Symmetry in the calculations is more striking if the pair of summation indices appears, one as a superscript, the other as a subscript. Therefore we allow the shifting of an index from the one position to the other at will. Thus,

$$x_i \equiv x^i.$$

Where no confusion will arise, indices may be placed outside of parentheses.

$$\left(\delta\delta - \frac{1}{a}\frac{1}{b}\right)_{ik}^{jl} = \delta_i^j \delta_k^l - \left(\frac{1_i^j}{a}\right)\left(\frac{1_k^l}{b}\right).$$

The standard notations for averages will be used.

$$\bar{x}_i = \left(\frac{1}{b}\right)^j x_{ij} = \left(\frac{1}{b}\right) \Sigma_j x_{ij} \tag{1}$$

$$\bar{x} = \left(\frac{1}{N}\right) \Sigma x_{ij,k} \quad . \tag{2}$$

Unless otherwise indicated, the symbol, Σ, will always stand for summation over *all* unrepeated indices including any already averaged under conventions (1) and (2). Thus,

$$\Sigma \bar{x}^2 = N\bar{x}^2.$$

The following simple formulas are fundamental to the arithmetic of this notation. They are obvious upon the expansion of the summations. Each index varies from 1 to a. These formulas are

$$\delta_i^j x_j = x_i,$$

$$\delta_i^j \delta_j^k = \delta_i^k,$$

$$\delta_i^j 1_j^k = 1_i^k,$$

$$1_i^j 1_j^k = a 1_i^k,$$

$$\delta_i^i = a,$$

$$x_i - \bar{x}_\cdot = \delta_i^j x_j - \left(\frac{1}{a}\right)_i^j x_j,$$

$$= \left(\delta - \frac{1}{a}\right)_i^j x_j.$$

The symbols of this notation obey the associative, commutative, and the distributive laws of simple arithmetic so that the operations of summation, multiplication, and squaring are very easy. Thus for the product of two linear forms we have

$$\bar{x}\bar{y} = \left(\frac{1}{a}\right)^i x_i \left(\frac{1}{b}\right)^j x_j = \left(\frac{1}{ab}\right)_j^i x_i x^j.$$

The sum of squares is obtained by the simple repetition of the form,

$$\Sigma x_i^2 = \Sigma(\delta_i^j x_j)^2 = (\delta_i^j x_j)(\delta_i^j x_j), \tag{3}$$
$$= (\delta_i^j x_j)(\delta_k^i x^k) = \delta_k^j x_j x^k.$$

Two other sums of squares occur so frequently that they should be particularly noted:

$$\Sigma \bar{x}^2 = \Sigma\left[\left(\frac{1}{a}\right)_i^j x_j\right]^2 = \left(\frac{1}{a}\right)_i^j x_j \left(\frac{1}{a}\right)_k^i x^k, \tag{4}$$
$$= \left(\frac{1}{a}\frac{1}{a}\right)_{ik}^{ij} x_j x^k = \left(\frac{1}{a}\right)_k^j \bar{x}_j x^k.$$

$$
\begin{aligned}
\Sigma(x_i - \bar{x})^2 &= \Sigma\left[\left(\delta - \frac{1}{a}\right)^i_s x_j\right]^2, \\
&= \left(\delta - \frac{1}{a}\right)^j_t \left(\delta - \frac{1}{a}\right)^i_k x_j x^k, \\
&= \left(\delta\delta - \frac{1}{a}\,\delta - \delta\,\frac{1}{a} + \frac{1}{a}\,\frac{1}{a}\right)^{ji}_{sk} x_j x^k, \qquad (5) \\
&= \left(\delta - \frac{1}{a} - \frac{1}{a} + \frac{1}{a}\right)^j_k x_j x^k, \\
&= \left(\delta - \frac{1}{a}\right)^j_k x_j x^k.
\end{aligned}
$$

The striking similarity in the coefficients of the second and final expressions for the summations in (3), (4), and (5) should not be overlooked.

Where we have multiple classification of the variables, we may operate on each index separately. For example, in a four-way analysis of variance we may have the quadratic form,

$$
\begin{aligned}
Q &= \Sigma\{\bar{x}_{ijk\cdot} - \bar{x}_{ij\cdot\cdot} - \bar{x}_{i\cdot k\cdot} + \bar{x}_{i\cdot\cdot\cdot}\}^2, \\
&= \Sigma\left\{\left[\delta\left(\delta\delta - \delta\,\frac{1}{c} - \frac{1}{b}\,\delta + \frac{1}{b}\,\frac{1}{c}\right)\frac{1}{d}\right]^{mnop}_{ijkl} x_{mnop}\right\}^2, \\
&= \Sigma\left\{\left[\delta\left(\delta - \frac{1}{b}\right)\left(\delta - \frac{1}{c}\right)\frac{1}{d}\right]^{mnop}_{ijkl} x_{mnop}\right\}^2, \\
&= \delta^m_q \left(\delta - \frac{1}{b}\right)^n_r \left(\delta - \frac{1}{c}\right)^o_s \left(\frac{1}{d}\right)^p_t x_{mnop} x^{qrst}.
\end{aligned}
$$

The rank is one of the important properties of a quadratic form or matrix. An experienced mathematician usually has a rule of thumb for determining the ranks of those quadratic forms occurring in statistical analysis. In order to formulate such rules of thumb into a simple and rigorous algebra, the author here defines a type of matrix multiplication which he calls "uncontracted matrix multiplication" and which he represents by the symbol, $\odot$.

Let $A = \| \alpha^j_i \|$ and $B = \| \beta^j_i \|$ be two matrices of any finite orders and with ranks R_A and R_B. We define the uncontracted product, $A \odot B$, as follows:

$$
\begin{aligned}
C &= A \odot B \\
&= \| \alpha^j_i \| \odot B \\
&= \| \alpha^j_i B \| \\
&= \begin{Vmatrix} \alpha^1_1 B & \alpha^2_1 B & \cdots \\ \alpha^1_2 B & \alpha^2_2 B & \cdots \\ \cdots & \cdots & \cdots \end{Vmatrix},
\end{aligned}
$$

where

$$\alpha_i^j L = \begin{Vmatrix} \alpha_i^j \beta_1^1 & \alpha_i^j \beta_1^2 & \cdots \\ \alpha_i^j \beta_2^1 & \alpha_i^j \beta_2^2 & \cdots \\ \cdots & \cdots & \cdots \end{Vmatrix}.$$

Thus the elements of C are

$$\gamma_{ik}^{jl} \equiv \alpha_i^j \beta_k^l.$$

We therefore see that whenever we have a matrix whose elements can be factored in the above manner, then the matrix can be expressed as the uncontracted product of simple matrices. Thus,

if $$\| \gamma_{ij}^{mn\cdots} \| \equiv \| (\alpha_i^m \beta_j^n \cdots) \|$$

then $$\| \gamma_{ij}^{mn\cdots} \| = \| \alpha_i^m \| \odot \| \beta_j^n \| \odot \cdots.$$

We shall now prove that the rank of the uncontracted product, $C = A \odot B$, of two matrices is equal to the product of the ranks. This follows because for the matrix, A, there always exists a set of elementary transformations defined by the equations,

$$T_A: \qquad {}_A\delta_i^j = \binom{j}{s}\binom{r}{i} \theta_n^s \theta_r^m \alpha_m^n, \qquad \theta_i^j,\ \theta_i^s \neq 0,\ i = j,$$

where the θ_i^j's, $i = j$, are coefficients providing for the multiplication of the elements of a row by a constant not zero; the θ_i^j's, $i \neq j$, are coefficients providing for the addition to the elements of a row a linear function of the corresponding elements of the other rows; the θ's are similar coefficients referring to columns; the symbol $\binom{j}{i}$ is an operator indicating the interchange of the ith and jth rows (columns); and the ${}_A\delta$'s have the values,

$$\begin{aligned} {}_A\delta_i^j &= 1, && i = j \leq R_A, \\ &= 0, && \text{otherwise.} \end{aligned}$$

This set of transformations reduces A to a diagonal matrix with R_A non-zero elements. A similar set of transformations,

$$T_B: \qquad {}_B\delta_k^l = \binom{l}{s}\binom{r}{k} \varphi_n^s \varphi_r^m \beta_m^n,$$

exists for the matrix B. We next define two sets of transformations by the equations,

$$T'_A: \qquad ({}_A\delta_i^j\ \beta_k^l) = \binom{jl}{sl}\binom{rk}{ik} \theta_n^s \theta_r^m (\alpha_m^n\ \beta_k^l),$$

$$T'_B: \qquad ({}_A\delta_i^j\ {}_B\delta_k^l) = \binom{jl}{js}\binom{ir}{ik} \varphi_n^s \varphi_r^m ({}_A\delta_i^j\ \beta_m^n),$$

which are also elementary because of their relationship to T_A and T_B. Now if we subject the matrix, $C = || (\alpha_i^j \beta_k^l) ||$ to the transformations T_A' followed by the transformations T_B', it will be reduced to the diagonal form $C = || ({}_A\delta_{iB}^j \delta_k^l) ||$ with exactly $R_A R_B$ non-zero elements. Therefore, since the rank of a matrix is invariant under elementary transformations, the rank of $C = A \odot B$ must be $R_A R_B$.

We shall now determine the ranks of several matrices which occur frequently in statistics:

$$A_1 = || 1_i || = || 1, 1, 1, \cdots ||, \quad R_1 = 1.$$

$$A_2 = || 1_i^j || = || 1_i \cdot 1^j || = || 1_i || \odot || 1^j ||,$$

$$R_2 = 1 \cdot 1 = 1.$$

$$A_3 = || \delta_i^j ||, \quad R_3 = a.$$

$$A_4 = \left\| \left(\delta - \frac{1}{a} \right)_i^j \right\|, \quad R_4 = a - 1.$$

The proof that $R_4 = a - 1$ involves two steps. First summing the rows of A_4 we have,

$$1^i \alpha_i^j = 1^i \delta_i^j - 1^i \left(\frac{1}{a} \right)^j = 1^j - \left(\frac{a}{a} \right)^j = 0$$

so that $R_4 \leq a - 1$. Second if we subtract the elements of the first row from the corresponding elements of each of the other rows we obtain,

$$A_4 \sim \left\| \begin{array}{c|c} 1 - \frac{1}{a} & \frac{1}{a} \\ \hline -1 & \delta_i^j \end{array} \right\| \begin{array}{l} i = 1 \\ i \neq 1. \end{array}$$

Since the $(a - 1)$st order determinant in the lower right-hand corner is not equal to zero, $R_4 \geq a - 1$.

Applying our theorem on uncontracted products, the ranks of complicated matrices can often be determined by inspection. Thus:

$$A_5 = \left\| \delta_i^j \left(\delta - \frac{1}{b} \right)_k^l \right\| = || \delta_i^j || \odot \left\| \left(\delta - \frac{1}{b} \right)_k^l \right\|,$$

$$R_5 = a \cdot (b - 1).$$

$$A_6 = \left\| \left(\delta - \frac{1}{a} \right)_i^j \left(\delta - \frac{1}{b} \right)_k^l \right\|,$$

$$R_6 = (a - 1)(b - 1).$$

$$A_7 = \left\| \left(\delta - \frac{1}{a} \right)_r^s \left(\delta - \frac{1}{a} \right)_t^j y_s y^t \right\| = \left\| \left[\left(\delta - \frac{1}{a} \right)^s y_s \right]_r \left[\left(\delta - \frac{1}{a} \right)_t y^t \right]^j \right\|,$$

$$R_7 = 1 \cdot 1 = 1.$$

The Matrix A_7 may be confusing at first sight. Note that each element, α_i^j, is a quadratic form in the y's. This form is of rank 1 and can be factored into two linear factors, one independent of j, the other independent of i.

To illustrate the application of these techniques to a fairly complicated problem, we shall construct and verify a design for the analysis of variance involving a regression line. It is known that sufficient conditions for such a design to be valid are:

1. The sum of the quadratic forms be equal to the sum of the squares of the variables, and
2. The sum of the ranks of the forms be equal to the number of variables.

We shall use the first condition to set up our design. Thus,

$$
\begin{aligned}
\Sigma x_{ij}^2 &= [\delta\delta]_{ij}^{kl} x_{kl} x^{ij}, \\
&\equiv \left\{ \left[\delta\delta - \delta \frac{1}{b} - \frac{1}{a}\delta + \frac{1}{a}\frac{1}{b} \right]_{ij}^{kl} + \left[\frac{1}{a}\frac{1}{b} \right]_{ij}^{kl} \right. \\
&\quad + \left[\frac{1}{a}\delta - \frac{1}{a}\frac{1}{b} \right]_{ij}^{kl} + \left[\left(\delta - \frac{1}{a} \right)_s^k \left(\delta - \frac{1}{a} \right)_i^t y^s y_t \left(\frac{1}{a\sigma_y^2} \right) \left(\frac{1}{b} \right)_j^l \right] \\
&\quad \left. + \left[\left(\delta \frac{1}{b} - \frac{1}{a}\frac{1}{b} \right)_{ij}^{kl} - \left(\delta - \frac{1}{a} \right)_s^k \left(\delta - \frac{1}{a} \right)_i^t y^s y_t \left(\frac{1}{a\sigma_y^2} \right) \left(\frac{1}{b} \right)_j^l \right] \right\} x_{kl} x^{ij}.
\end{aligned} \tag{6}
$$

Rewriting this in the usual notation, we have for our tenative design,

$$
\begin{aligned}
\Sigma x_{ij}^2 &= \Sigma [x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}]^2 + \Sigma [\bar{x}]^2 + \Sigma [\bar{x}_{.j} - \bar{x}]^2 \\
&\quad + \Sigma [(r\sigma_x/\sigma_y)(y_i - \bar{y})]^2 + \Sigma [(\bar{x}_{i.} - \bar{x}) - (r\sigma_x/\sigma_y)(y_i - \bar{y})]^2.
\end{aligned} \tag{7}
$$

In order to determine the corresponding equation for the ranks, we rewrite (6) in the form,

$$
\begin{aligned}
\Sigma x_{ij}^2 &= \left\{ \left(\delta - \frac{1}{a} \right)_i^k \left(\delta - \frac{1}{b} \right)_j^l + \left(\frac{1}{a} \right)_i^k \left(\frac{1}{b} \right)_j^l + \left(\frac{1}{a} \right)_i^k \left(\delta - \frac{1}{b} \right)_j^l \right. \\
&\quad + \left[\left(\delta - \frac{1}{a} \right)_s^k y^s \right] \left[\left(\delta - \frac{1}{a} \right)_i^t y_t \right] \left(\frac{1}{a\sigma_y^2} \right) \left(\frac{1}{b} \right)_j^l \\
&\quad \left. + \left[\left(\delta - \frac{1}{a} \right)_i^k - \left(\delta - \frac{1}{a} \right)_s^k \left(\delta - \frac{1}{a} \right)_i^t y^s y_t \left(\frac{1}{a\sigma_y^2} \right) \right] \left(\frac{1}{b} \right)_j^l \right\} x_{kl} x^{ij}.
\end{aligned} \tag{8}
$$

First we must determine the rank of the unfamiliar matrix,

$$
A_8 = \left\| \left(\delta - \frac{1}{a} \right)_i^j - \left(\delta - \frac{1}{a} \right)_s^j \left(\delta - \frac{1}{a} \right)_i^t y^s y_t / a\sigma_y^2 \right\|.
$$

We see that the rank of A_8 cannot be greater than $a - 2$ because two linear relations exist between the rows, namely,

$$
1^t \alpha_i^j = 0, \quad \text{since} \quad 1^t \left(\delta - \frac{1}{a} \right)_i^t = 0,
$$

$$
y^t \alpha_i^j = 0, \quad \text{since} \quad \left(\delta - \frac{1}{a} \right)_i^t y_t y^t = a\sigma_y^2.
$$

To show that the rank of A_8 cannot be less than $a-2$, we subtract the elements of the first row from the corresponding elements of each of the last $a-2$ rows, giving,

$$A_8 \sim \left\| \begin{array}{c|c|c} \dfrac{\alpha_i^1}{\alpha_i^1 - \alpha_1^1} & \dfrac{\alpha_i^2}{-\left(\delta - \dfrac{1}{a}\right)_s^2 y^s(\delta_i^t - \delta_1^t)y_t} & \dfrac{\alpha_i^j}{\delta_i^j - \dfrac{\left(\delta - \dfrac{1}{a}\right)_s^j y^s(\delta_i^t - \delta_1^t)y_t}{a\sigma_y^2}} \end{array} \right\| \begin{array}{l} i = 1, 2 \\ i \neq 1, 2 \end{array}$$

Multiplying each element of the second column by $-\left(\delta - \dfrac{1}{a}\right)_s^j y^s \Big/ \left(\delta - \dfrac{1}{a}\right)_s^2 y^s$ and adding the result to the corresponding element of the jth column for $j = 3, 4, \cdots a$, we see that the $(a-2)$th order determinant in the lower right-hand corner becomes $|\delta_i^j|$ which is not equal to zero. Therefore the rank of A_8 must equal $a-2$.

Referring to equation (8), we now write down the corresponding equation for ranks using the theorem on uncontracted products. Thus,

$$\Sigma \text{ Ranks} = (a-1)(b-1) + (1)(1) + (1)(b-1) + (1)(1)(1) + (a-2)(1),$$

$$= ab.$$

Hence the quadratic forms in the right member of equation (7) are mutually independent and each, measured in units of the variance of the population, is distributed as is Chi-square with the appropriate number of degrees of freedom.

A SYMMETRIC METHOD OF OBTAINING UNBIASED ESTIMATES AND EXPECTED VALUES

By Paul L. Dressel

Michigan State College, East Lansing, Michigan

The problem of finding the relationship between moment functions of a sample and moment functions of the population from which the sample was obtained has, of necessity, received much attention. The problem has two parts. first, to find the expected value of a given sample moment function; second, to find the estimate of a given population moment function. Thus, if m_i represent the ith central moment of a sample and μ_i represent the ith central moment of the population, the first part of the problem requires that we find the mean value of m_i for all possible samples of a given size and express it in term of the μ_i's The second part requires that we find a function of the m_i's such that the mean value, taken for all possible samples of a given size, be a given μ_i. For the case $i = 4$ we have the well known results:

$$E[m_4] = \frac{(n-1)(n^2-3n+3)}{n^3}\mu_4 + \frac{3(n-1)(2n-3)}{n^3}\mu_2^2,$$

$$E^{-1}[\mu_4] = \frac{n^2(n^2-2n+3)}{n^{(4)}}m_4 - \frac{3n^2(2n-3)}{n^{(4)}}m_2^2.$$

These results are based on the assumption of an infinite population. In spite of the inverse relationship existing between estimates and expected value, the expressions above show no simple relationship. This lack of simplicity of relationship between estimate and expected value is directly traceable to the fact that such results are usually obtained for infinite populations. When results are obtained for finite populations a symmetry is found to exist which reduces to a single problem the two parts stated above. Since this should be evident to anyone upon reflection, the main purpose of the present paper may be considered as that of indicating one method of demonstrating the result stated above as well as showing relationship of this method to material appearing in previously published papers

Consider a finite population consisting of N items $x_1 \cdots x_N$ and samples of n items taken from that population, the sampling being done without replacement. We shall utilize the power product notation of P. S. Dwyer [1; p. 13]

$$(1) \qquad (q_1 \cdots q_r) = \sum_{i_1 \neq i_2 \neq \cdots \neq i_r}^{n} x_{i_1}^{q_1} x_{i_2}^{q_2} \cdots x_{i_r}^{q_r}$$

to represent a power product formed for the sample and

$$[q_1 \cdots q_r] = \sum_{i_1 \neq i_2 \neq \cdots \neq i_r}^{N} x_{i_1}^{q_1} x_{i_2}^{q_2} \cdots x_{i_r}^{q_r} \tag{2}$$

to represent like power products formed for the population. An arbitrary moment function of weight r of the sample is indicated by

$$\Sigma a_{q_1^{\pi_1} \cdots q_t^{\pi_t}} \frac{r!}{(q_1!)^{\pi_1} \cdots (q_t!)^{\pi_t} \pi_1! \cdots \pi_t!} (q_1)^{\pi_1} \cdots (q_t)^{\pi_t} \tag{3}$$

and likewise a moment function of the population is indicated by

$$\Sigma A_{q_1^{\pi_1} \cdots q_t^{\pi_t}} \frac{r!}{(q_1!)^{\pi_1} \cdots (q_t!)^{\pi_t} \pi_1! \cdots \pi_t!} [q_1]^{\pi_1} \cdots [q_t]^{\pi_t} \tag{4}$$

where the summation extends over all partitions of r.

It now is convenient to express each of the expressions (3) and (4) in terms of power products. We shall utilize for this purpose an expansion theorem which is the converse of a theorem stated by Dwyer, [1; p. 34] and [2; pp. 37–39], which can be proved in a similar fashion.

This converse theorem follows:

If any isobaric sum of products of power sums indicated by

$$\Sigma A_{q_1^{\pi_1} \cdots q_t^{\pi_t}} \frac{r!}{(q_1!)^{\pi_1} \cdots (q_t!)^{\pi_t} \pi_1! \cdots \pi_t!} [q_1]^{\pi_1} [q_2]^{\pi_2} \cdots [q_t]^{\pi_t} \tag{5}$$

be expanded in terms of power products in a form indicated by

$$\Sigma B_{p_1^{\pi_1} \cdots p_s^{\pi_s}} \frac{r!}{(p_1!)^{\pi_1} \cdots (p_s!)^{\pi_s} \pi_1! \cdots \pi_s!} [p_1^{\pi_1} \cdots p_s^{\pi_s}] \tag{6}$$

then the coefficient B_r of the power sum $[r]$ is given by

$$B_r = \Sigma \frac{r!}{(p_1!)^{\pi_1} \cdots (p_s!)^{\pi_s} \pi_1! \cdots \pi_s!} A_{p_1^{\pi_1} \cdots p_s^{\pi_s}} \tag{7}$$

and the coefficient $B_{r_1 \cdots r_m}$ of $[r_1 r_2 \cdots r_m]$ is

$$B_{r_1 r_2 \cdots r_m} = \overline{B_{r_1} B_{r_2} \cdots B_{r_m}} \tag{8}$$

where the barred product indicates a symbolic multiplication by suffixing of subscripts.

This is exemplified by

$$\begin{aligned} B_{32} = \overline{B_3 B_2} &= \overline{(A_3 + 3A_{21} + A_{111})(A_2 + A_{11})} \\ &= A_{32} + A_{311} + 3A_{221} + 4A_{2111} + A_{11111}. \end{aligned}$$

Using this theorem the moment functions (3) and (4) are easily expanded in terms of power products. In this latter form the expected value of the sample moment function is easily found by utilizing the fact that

$$E\left(\frac{(q_1 \cdots q_s)}{n^{(s)}}\right) = \frac{[q_1 \cdots q_s]}{N^{(s)}}.$$

Now if the expected value of the sample moment function be equated to the population moment function (both being in power product form) we obtain a set of equations connecting the coefficients of a sample moment function and a population moment function. Since either the coefficients of the sample moment function or those of the population moment function may be assigned and the others solved for, this set of equations enables one to solve two problems. First, we may find unbiased estimates—moment functions of the sample such that their expected value is some preassigned population moment function. Second, we may find expected values—moment functions of the population such that they are expected values of some preassigned sample moment function. From the symmetry of this set of equations, we shall see that any result obtained from the system has, through the symmetry, a dual role.

The foregoing discussion may be clarified by an example. Let $A_2[2] + A_{11}[1]^2$ be the population moment function. In terms of power products this becomes $(A_2 + A_{11})[2] + A_{11}[11]$. The sample moment function $a_2(2) + a_{11}(1)^2$ becomes in terms of power products $(a_2 + a_{11})(2) + a_{11}(11)$ and its expected value is

$$\frac{n}{N}(a_2 + a_{11})[2] + \frac{n^{(2)}}{N^{(2)}} a_{11}[11].$$

By equating this to the population moment function above we obtain

$$n^{(2)}a_{11} = N^{(2)}A_{11},$$
$$n(a_2 + a_{11}) = N(A_2 + A_{11}),$$

and the symmetry of the system is apparent.
If

$$\rho_i = \frac{n^{(i)}}{N^{(i)}}, \qquad \tau_i = \frac{N^{(i)}}{n^{(i)}} = \frac{1}{\rho_i},$$

the solutions of the system are

$$(9) \qquad \begin{aligned} a_{11} &= \tau_2 A_{11}, & A_{11} &= \rho_2 a_{11}, \\ a_2 &= \tau_1 A_2 + (\tau_1 - \tau_2)A_{11}, & A_2 &= \rho_1 a_2 + (\rho_1 - \rho_2)a_{11}. \end{aligned}$$

In a similar manner if we use moment functions of weight 3 we begin with

$$A_3[3] + 3A_{21}[2][1] + A_{111}[1]^3,$$
$$a_3(3) + 3a_{21}(2)(1) + a_{111}(1)^3,$$

and obtain the system of equations

$$n^{(3)}a_{111} = N^{(3)}A_{111}$$
$$n^{(2)}(a_{21} + a_{111}) = N^{(2)}(A_{21} + A_{111})$$
$$n(a_3 + 3a_{21} + a_{111}) = N(A_3 + 3A_{21} + A_{111})$$

with solutions

$$
\begin{aligned}
A_{111} &= \rho_3 a_{111}\,, \\
A_{21} &= \rho_2 a_{21} + (\rho_2 - \rho_3) a_{111}\,, \\
A_3 &= \rho_1 a_3 + 3(\rho_1 - \rho_2) a_{21} + (\rho_1 - 3\rho_2 + 2\rho_3) a_{111}\,.
\end{aligned} \tag{10}
$$

The solutions for the a's in terms of the A's are obtainable from the given results in an obvious manner.

If we use the Carver functions [3; p. 104]

$$
\begin{aligned}
P_1 &= \rho_1\,, & P_{11} &= \rho_2 & \cdots \quad P_{1^i} &= \rho_i\,, \\
P_2 &= \rho_1 - \rho_2\,, & P_{21} &= \rho_2 - \rho_3 & \cdots & \\
P_3 &= \rho_1 - 3\rho_2 + 2\rho_3\,, & P_{22} &= \rho_2 - 2\rho_3 + \rho_4 \cdots & & \\
P_4 &= \rho_1 - 7\rho_2 + 12\rho_3 - 6\rho_4\,, & & & & \\
&\vdots & & & &
\end{aligned}
$$

or in general

$$
P_r = \sum_{i=1}^{r} \rho_i \sum (-1)^{i-1} \frac{r!(i-1)!}{(p_1!)^{\pi_1} \cdots (p_s!)^{\pi_s} \pi_1! \cdots \pi_s!} \tag{11}
$$

and

$$
P_{r_1 r_2 \cdots r_s} = \overline{\overline{P_{r_1} P_{r_2} \cdots P_{r_s}}}
$$

where the double barred product indicates a symbolic multiplication by addition of subscripts exemplified by

$$
\begin{aligned}
P_{32} = \overline{\overline{P_3 P_2}} &= (\rho_1 - 3\rho_2 + 2\rho_3)(\rho_1 - \rho_2) \\
&= \rho_2 - 4\rho_3 + 5\rho_4 - 2\rho_5\,;
\end{aligned}
$$

the results (9) and (10) may be written

$$
\begin{aligned}
A_{11} &= P_{11} a_{11}\,, & A_3 &= P_1 a_3 + 3P_2 a_{21} + P_3 a_{111}\,, \\
A_2 &= P_1 a_2 + P_2 a_{11}\,, & A_{21} &= P_{11} a_{21} + P_{21} a_{111}\,, \\
& & A_{111} &= P_{111} a_{111}\,.
\end{aligned}
$$

Similarly for weight 4 we obtain

$$
\begin{aligned}
A_4 &= P_1 a_4 + 4P_2 a_{31} + 3P_2 a_{22} + 6P_3 a_{211} + P_4 a_{1111}\,, \\
A_{31} &= P_{11} a_{31} + 3P_{21} a_{211} + P_{31} a_{1111}\,, \\
A_{22} &= P_{11} a_{22} + 2P_{21} a_{211} + P_{22} a_{1111}\,, \\
A_{211} &= P_{111} a_{211} + P_{211} a_{1111}\,, \\
A_{1111} &= P_{1111} a_{1111}\,.
\end{aligned}
$$

In general

$$A_r = \Sigma P_{\pi_1+\pi_2+\cdots+\pi_s} \frac{r!}{(p_1!)^{\pi_1}(p_2!)^{\pi_2} \cdots (p_s!)^{\pi_s}\pi_1! \cdots \pi_s!} a_{p_1^{\pi_1}p_2^{\pi_2}\cdots p_s^{\pi_s}}, \tag{12}$$

and

$$A_{r_1r_2\cdots r_m} = \overline{A_{r_1}A_{r_2} \cdots A_{r_m}}, \tag{13}$$

where as before the barred product indicates a symbolic multiplication by suffixing of subscripts.

If in

$$\Sigma a_{q_1^{\pi_1}\cdots q_t^{\pi_t}} \frac{r!}{(q_1!)^{\pi_1} \cdots (q_t!)^{\pi_t}\pi_1! \cdots \pi_t!} (q_1)^{\pi_1} \cdots (q_t)^{\pi_t} \tag{14}$$

$$a_{q_1^{\pi_1}\cdots q_t^{\pi_t}} = \frac{(-1)^{\pi_1+\pi_2+\cdots+\pi_t}(\pi_1 + \pi_2 + \cdots + \pi_t - 1)!}{n^{\pi_1+\pi_2+\cdots+\pi_t}} \tag{15}$$

the moment function of the sample which is thereby represented is the Thiele seminvariant l_r of the sample. If the A's are solved for by means of the appropriate set of equations the expected value of l_r is found. Thus we find

$$\begin{aligned}
E[l_2] &= \frac{N^2 n^{(2)}}{N^{(2)} n^2}\lambda_2, \\
E[l_3] &= \frac{N^3 n^{(3)}}{N^{(3)} n^3}\lambda_3, \\
E[l_4] &= \frac{N^4 n^{(4)}}{N^{(4)} n^4}\lambda_4 + \frac{n^{(2)} N^2}{n^4 N^{(4)}}(n - N)(Nn - 6)\kappa_4, \\
E[l_2^2] &= \frac{N^4 n^{(4)}}{N^{(4)} n^4}\lambda_2^2 - \frac{N^2 n^{(2)}}{N^{(4)} n^4}(n - N)(nN - n - N - 1)\kappa_4, \\
E[l_5] &= \frac{N^5 n^{(5)}}{N^{(5)} n^5}\lambda^5 + \frac{5N^3 n^{(3)}}{N^{(5)} n^5}(n - N)(Nn - 12)\kappa_5, \\
E[l_3 l_2] &= \frac{N^5 n^{(5)}}{N^{(5)} n^5}\lambda_3\lambda_2 - \frac{N^3 n^{(3)}}{N^{(5)} n^5}(n - N)(Nn - n - N - 5)\kappa_5,
\end{aligned} \tag{16}$$

where the κ system of seminvariants used here is defined by

$$\begin{aligned}
\kappa_{2r} &= \frac{1}{2}\sum_{i=0}^{2r}(-1)^i \binom{2r}{i}\mu_i\mu_{2r-i}, \\
\kappa_{2r+1} &= \sum_{i=0}^{r}(-1)^{i+r}\binom{2r}{i+r}\frac{2i+1}{i+r+1}\mu_{r-i}\mu_{r+i+1}.
\end{aligned} \tag{17}$$

By virtue of the symmetry noted earlier it follows that the estimates of the Thiele seminvariants and products of these seminvariants of weight ≤ 5 are

obtainable from the last results by replacing E by E^{-1} (estimate of), κ_i by k_i, l_i by λ_i, and N by n. In this manner we find that L_4, the estimate of λ_4 is

$$L_4 = E^{-1}[\lambda_4] = \frac{n^4 N^{(4)}}{n^{(4)} N^4} l_4 + \frac{N^{(2)} n^2}{N^4 n^{(4)}} (N - n)(Nn - 6)k_4. \tag{18}$$

It is of some interest to note in the results (16) above that in those expected values or estimates which contain more than one term the factor $N - n$ occurs in the second term. This, and the form of other coefficients involved in the terms, shows that as the sample size approaches the population size the sample seminvariants approach the population seminvariants. Another characteristic of such results as those given in (16) is that infinite sampling formulas are easily obtainable therefrom. Thus if in L_4 given in (18) $N \to \infty$, we find

$$\begin{aligned} L_4 &= \frac{n^4}{n^{(4)}} l_4 + \frac{n^3}{n^{(4)}} k_4 \\ &= \frac{n^3(n+1)}{n^{(4)}} m_4 - \frac{3n^3(n-1)}{n^{(4)}} m_2^2, \end{aligned}$$

the first of these forms checking the result given by Dressel [4; p. 45] and the second form being identical with that given by Fisher [5].

The results exhibited above for finite sampling may lead to a mistaken idea about the simplicity of the results. Simplicity decreases rapidly as the weight increases. Thus for weight 6 we find

$$\begin{aligned} E[l_6] = {} & \frac{N^6 n^{(6)}}{N^{(6)} n^6} \lambda_6 + \frac{2N^4 n^{(4)}}{N^{(6)} n^5} (n - N)(Nn - 20)[8\mu_6 - 15\mu_4\mu_2 + 10\mu_3^2 - 45\mu_2^3] \\ & + \frac{N^3 n^{(3)}}{N^{(6)} n^5} (n - N)[Nn(n + N) - 12nN + 60] \\ & \cdot[11\mu_6 + 105\mu_4\mu_2 - 50\,\mu_3^2 + 60\mu_2^3] \\ & - \left\{ \frac{4N^2 n^{(2)}}{N^{(6)} n^6} (n - N)[Nn(N^2 + nN + n^2) - 14nN(N + n) + 71Nn - 120] \right. \\ & \left. + \frac{10Nn^{(3)}}{N^{(4)} n^5} (n - N) + \frac{6n^{(2)}}{N^{(4)} n^5} (n - N)(N + n - 5) - \frac{2n^{(2)}}{N^{(3)} n^5} (n - N) \right\} \kappa_6. \end{aligned} \tag{19}$$

Again by letting $N \to \infty$ infinite sampling results are obtained. Much of this last result vanishes in that case.

It has been demonstrated that the κ system of seminvariants are invariant under estimation in the case of infinite sampling [4; p. 53]. It is therefore of some interest to note that this system also possesses the property for finite sampling without replacement The proof of this is quite simple. Denote the estimate of κ_i by K_i and the fundamental relations are

$$K_{2r} = \frac{n^2}{n^{(2)}} k_{2r}, \qquad K_{2r+1} = \frac{n^3}{n^{(3)}} k_{2r+1}.$$

These expressions hold for any n and hence for a population of N. Let K'_{2r} and K'_{2r+1} denote functions corresponding to K_{2r} and K_{2r+1} but with population moments replacing sample moments and we have

$$K'_{2r} = \frac{N^2}{N^{(2)}} \kappa_{2r}, \qquad K'_{2r+1} = \frac{N^3}{N^{(3)}} \kappa_{2r+1}.$$

Since the power product mode of formulation of K_{2r} and K_{2r+1} insures that

$$E[K_{2r}] = K'_{2r}, \qquad E[K_{2r+1}] = K'_{2r+1}$$

it follows that

$$E[K_{2r}] = E\left[\frac{n^2}{n^{(2)}} k_{2r}\right] = K'_{2r} = \frac{N^2}{N^{(2)}} \kappa_{2r},$$

or

$$E[k_{2r}] = \frac{n^{(2)} N^2}{n^2 N^{(2)}} \kappa_{2r}.$$

Similarly

$$E[k_{2r+1}] = \frac{n^{(3)} N^3}{n^3 N^{(3)}} \kappa_{2r+1},$$

thus establishing the theorem stated above.

REFERENCES

[1] P. S. Dwyer, "Combined expansions of products of symmetric power sums and sums of symmetric power products with applications to sampling," Part I, *Annals of Math. Stat.*, Vol. 9 (1938), pp 1-47. Part II, Vol. 9, (1938) pp. 97-132.

[2] P. S. Dwyer, "Moments of any rational integral isobaric sample moment function," *Annals of Math. Stat.*, Vol 8 (1937), pp 21-65.

[3] H C. Carver, "Fundamentals of the theory of sampling," *Annals of Math. Stat.*, Vol. 1 (1930), pp. 101-121; 260-274.

[4] P. L. Dressel, "Seminvariants and their estimates," *Annals of Math. Stat.*, Vol. 11 (1940), pp 33-57.

[5] R. A. Fisher, "Moments and product moments of sampling distributions," *Proc. Lond. Math. Soc.*, Vol. 2 (30), (1929), pp. 199-238.

DETERMINATION OF SAMPLE SIZES FOR SETTING TOLERANCE LIMITS

By S. S Wilks

Princeton University, Princeton, N. J.

1. Introduction. In the mass production of a given product or apparatus piece-part, Shewhart[1] has discussed a practical procedure for detecting the existence of assignable causes of variation in a given quality characteristic of the product as measured by a variable x For example, x may be the thickness in inches of a washer or the tensile strength in pounds of a small aluminum casting made according to a given set of specifications; x varies in value from washer to washer or from casting to casting. Now suppose assignable causes of variability in x have been detected by Shewhart's procedure and have been sufficiently well eliminated by making appropriate refinements in the manufacturing process so that for all practical purposes the remaining variability may be considered "random," thus allowing us to assume that we have a statistical universe U in which x is a random variable with some distribution law $f(x)$ $f(x)$ is, in general, unknown and cannot be determined until long after the refined manufacturing operation has been under way. Two types of situations arise in practice, one in which x is a discrete variable taking on only certain isolated values as for example 1, 2, 3, $\cdots$, etc. with corresponding probabilities $p(1)$, $p(2)$, $\cdots$, the other being that in which x is essentially a continuous variable over some range with a corresponding probability density function $f(x)$. In this paper we shall consider the latter type of variable.

The problem now arises as to how we should calculate a tolerance range (L_1, L_2) for x from a sample, and how large the sample should be in order for the tolerance range to have a given degree of stability. More specifically, *for a given method of calculating tolerance limits, how large should our sample be in order that the proportion P of the universe included between L_1 and L_2 have an average value a, and will be such that the probability is at least p that P will lie between two given numbers, say b and c?* For example, if a tolerance range is obtained by using a *truncated sample range*, that is by letting L_1 be the greatest of the r smallest values in a sample and L_2 the smallest of the r largest values, r being chosen so that $E(P) = .99$, how large should the sample size, say n, be in order for the probability to be .9 that P would lie between .985 and .995? A similar question can be asked when the setting of only one tolerance limit is under consideration.

[1] W. A. Shewhart, *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Company, New York, 1931.

2. Tolerance ranges from truncated sample ranges. Suppose that nothing is known about the distribution function $f(x)$ except enough to enable us to assume that it is continuous. Let a be the average value which P is to have, and suppose a sample of size n is drawn from the universe P so that $[(1-a)(n+1)]/2 = r$, say, is a positive integer. Let $x_1, x_2, \cdots, x_n$ be the sample values of x arranged in order of increasing magnitude. Let $L_1 = x_r$ and $L_2 = x_{n-r+1}$. The distribution law, say $g(P)$ of P the proportion of the universe included between these values of L_1 and L_2 is given by

$$(1) \quad g(P)\,dP = \frac{\Gamma(n+1)}{\Gamma[a(n+1)]\Gamma[(1-a)(n+1)]} P^{a(n+1)-1}(1-P)^{(1-a)(n+1)-1}\,dP.$$

This follows at once from the joint distribution law of x_r and x_{n-r+1} which can be derived as follows: Consider the x axis as being divided into k mutually exclusive intervals $I_1, I_2, \cdots, I_k$ with $p_1, p_2, \cdots, p_k$ as the associated probabilities $\left(\sum_1^k p_i = 1\right)$. In a sample of size n the probability that $n_1, n_2, \cdots, n_k$ $\left(\sum_1^k n_i = n\right)$ values of x will fall into $I_1, I_2, \cdots, I_k$ respectively is given by the well-known multinomial distribution law

$$(2) \qquad \frac{n!}{n_1!n_2!\cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}.$$

To get the distribution of x_r and x_{n-r+1} we take $k = 5$ and for $I_1, I_2, \cdots, I_5$ we take the intervals $(-\infty, x_r)$, $(x_r, x_r + dx_r)$, $(x_r + dx_r, x_{n-r+1})$, $(x_{n-r+1}, x_{n-r+1} + dx_{n-r+1})$, $(x_{n-r+1} + dx_{n-r+1}, \infty)$ respectively. The values of $p_1, p_2, \cdots, p_5$ are the integrals of $f(x)\,dx$ over these five intervals respectively and the values of $n_1, n_2, \cdots, n_5$ are $r-1, 1, n-2r, 1, r-1$ respectively. By substituting these values of the p's and n's in (2) and neglecting terms of order higher than $dx_r\,dx_{n-r+1}$ the probability element for x_r and x_{n-r+1} is found at once to be[2]

$$(3) \quad \frac{n!}{[(r-1)!]^2(n-2r)!}\left(\int_{-\infty}^{x_r} f(x)\,dx\right)^{r-1}\left(\int_{x_{n-r+1}}^{\infty} f(x)\,dx\right)^{r-1} \cdot\left(\int_{x_r}^{x_{n-r+1}} f(x)\,dx\right)^{n-2r} f(x_r)f(x_{n-r+1})\,dx_r\,dx_{n-r+1}.$$

Now let $\int_{-\infty}^{x_r} f(x)\,dx = u$, $\int_{x_{n-r+1}}^{\infty} f(x)\,dx = v$, then since $du = f(x_r)\,dx_r$ and $dv = -f(x_{n-r+1})\,dx_{n-r+1}$, the probability element of u and v may be written as

$$(4) \qquad \frac{\Gamma(n+1)}{\Gamma^2(r)\Gamma(n-2r+1)} u^{r-1}v^{r-1}(1-u-v)^{n-2r}\,du\,dv,$$

[2] For a discussion and a rather complete bibliography of the probability theory of "extreme values" such as x_r and x_{n-r+1} see E. J. Gumbel, "Les valeurs extrêmes des distributions statistiques," *Annales de l'Institut H. Poincaré* (1935).

the region of u and v of non-zero probability being the triangle bounded by the u and v axes and the line $u + v = 1$. Making the change of variables $1 - u - v = P$ and $u = Q$, integrating with respect to Q, and setting $r = (1/2)(1 - a)(n + 1)$ we find the distribution of P, the proportion of the universe included between x_r and x_{n-r+1} to be (1). It should be remarked that even if L_1 and L_2 are obtained by asymmetrical truncation by taking $L_1 = x_s$, $L_2 = x_t$ where $t - s = n - 2r + 1$, the distribution of $P = \int_{x_s}^{x_t} f(x)\,dx$ remains unchanged. Thus for a given p, by taking $L_1 = x_s$ and $L_2 = x_t$ where $t - s = n - 2r + 1 = a(n + 1)$, and choosing the smallest value of n for which $\int_b^c g(P)\,dP \geq p$ and such that $(1 - a)(n + 1)$ is a positive integer we have provided the answer to the italicized question for one method of calculating L_1 and L_2; a method which is valid for any unknown continuous distribution $f(x)$.

As an example, suppose we take $a = .99$, $b = .985$, $c = .995$ and $p = .99$. The size of sample required is found to be 1000 (999 to be exact). In fact in this case the probability of P being between .985 and 995 is .992. In this example, we may therefore make the statement that if x is a continuous variable under statistical control, and if samples of size 1000 are taken, the tolerance limits L_1 and L_2 taken as the fifth smallest and fifth largest values of x in the sample respectively, will, on the average, include 99% of the universe between them and furthermore, the tolerance limits calculated in this way for samples of size 1000 will, in about 99.2% of the samples, include between 98.5% and 99.5% of the universe between them.

If L_1 and L_2 are taken as the smallest and largest values of x in the sample respectively (corresponding to $r = 1$, i.e. sample range with no truncation), then in samples of size 1000, these tolerance limits will, on the average include 99.8% of the universe between them and the probability is .996 that L_1 and L_2 will include at least 99.5% of the universe between them. If the largest and smallest values of x in samples are used as tolerance limits and if we wish to state that the probability is .99 that such tolerance limits will include at least 99% of the universe, the size of sample required is 660. If the probability is lowered to .95 of including at least 99% of the universe, with such tolerance limits, the size of sample required is 130. Engineering statisticians[3] have pointed out on basis of practical experience the need of using samples of 100 to 1000 on even more cases in order to set tolerance limits which will include at least 99% of the universe with a satisfactorily high degree of certainty. The examples we have given based on sizes 1000, 660 and 130 will indicate the degree of stability to be expected for tolerance ranges for samples in this range of sizes. The degree of stability of the tolerance limits for samples of the size range 500 to 1000 appears to be of about the order of that demanded by the engineering statistician.

---

[3]Cf. W. A. Shewhart, *Statistical Methods from the Point of View of Quality Control*, The Graduate School of the U.S. Department of Agriculture, Washington (1939). P. 63.

In some cases it may be desirable to determine the size of samples so as to control the tolerance limits L_1 and L_2 individually, that is so that the probability is at least p that the proportions of the universe contained in the tails of the distribution cut off by L_1 and L_2 are in both cases between two given numbers, say d and e. In this case we would determine the least value of n so that

$$\int_d^e \int_d^e h(u, v)\, du\, dv \geq p \tag{5}$$

where $h(u, v)\, du\, dv$ denotes the function given by (4). For example, suppose $p = .99$, $d = 0$, $e = .005$. $r = 1$. The size of the sample needed is 1060. Thus in samples of size 1060, the probability is .99 that L_1 and L_2 taken as the smallest and the largest values in the sample respectively will cut off tails of the universe such that each tail will include not more than 0.5% of the universe.

If it is desired to set only one tolerance limit, say L_1, then the distribution of u would be used. This can be found by integrating (4) with respect to v. The distribution is

$$\frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)}\, u^{r-1}(1-u)^{n-r}\, du. \tag{6}$$

The probability p that the proportion of the universe in the tail which will be cut off by L_1 is between d and e is given by integrating the expression (6) from d to e. The value of n required to obtain any given value of p can then be determined. For example, in the case where $p = .99$, $d = 0$, $e = .005$, $r = 1$, the size of the sample needed is 920.

3. Tolerance range for a normal universe. The method of setting tolerance limits discussed in Section 2 assumes nothing about the distribution $f(x)$ except that it is continuous. If $f(x)$ can be assumed to have a given functional form involving unknown parameters, methods based on the theory of statistical estimation and having greater efficiency than those already discussed could be used for setting tolerance limits. We shall not go into a general discussion of such methods here although it does appear desirable to consider one very important example of the application of the methods. Suppose $f(x)$ can be assumed to be a normal distribution function with unknown mean m and variance σ^2. In a sample of size n let $\bar{x}$ be the sample mean and let $s^2 = \sum_1^n (x_i - \bar{x})^2/(n-1)$. Let us consider as tolerance limits L_1' and L_2' the quantities $\bar{x} \pm ks$. The proportion P' of the universe included between these limits is

$$P' = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{\bar{x}-ks}^{\bar{x}+ks} e^{-\frac{1}{2}(x-m)^2/\sigma^2}\, dx. \tag{7}$$

We wish to determine k so that $E(P') = a$. It can be verified by straightforward analysis that $E(P')$, defined by $\int_{-\infty}^{\infty} \int_0^{\infty} P' f(\bar{x}, s)\, ds\, d\bar{x}$, has the value

$$\frac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\,\Gamma((n-1)/2)} \int_{-t}^{t} \frac{dx}{(1 + x^2/(n-1))^{n/2}}, \qquad t = k\sqrt{\frac{n}{n+1}} \tag{8}$$

where $f(\bar{x}, s)$ is the well-known distribution of $\bar{x}$ and s given by

$$(9) \qquad \frac{\sqrt{n}(n-1)^{(n-1)/2}s^{n-2}}{2^{n/2-1}\sigma^n\sqrt{\pi}\Gamma((n-1)/2)}\, e^{-\frac{1}{2}[n(\bar{x}-m)^2+(n-1)s^2]/\sigma^2}.$$

Therefore the tolerance limits L_1' and L_2' which will include, on the average, a proportion a of the universe between them are

$$(10) \qquad \bar{x} \pm t_a\sqrt{(n+1)/n}\cdot s$$

where t_a is the value of t for which the integral in (8) has the value a. The value of t_a can be found from Fisher's t-table for $n-1$ degrees of freedom, and for certain values of a including .99, .95, etc. and for values of n up to 30. Although the tolerance limits (10) will include, on the average, the proportion a of the universe between them, we must now investigate the size of sample needed to obtain a given degree of stability of P'. The exact distribution of P' seems to be too complicated to be of any practical value. It is not difficult to verify that to within terms of order $1/n$, the variance of P' is given by

$$(11) \qquad \sigma_{P'}^2 = t_a^2 e^{-t_a^2}/(\pi n).$$

The variance of P, the proportion of the universe included between x_r and x_{n-r+1}, to within terms of order $1/n$ is given by

$$(12) \qquad \sigma_P^2 = a(1-a)/n.$$

For a large sample of a given size, say $n = 100$ or more, a simple comparison of the stabilities of the two tolerance ranges (x_r, x_{n-r+1}) and $(\bar{x} \pm t_a\sqrt{(n+1)/n}\cdot s)$ can be made by comparing σ_P^2 and $\sigma_{P'}^2$. For $a = .99$, the efficiency ratio $\sigma_{P'}^2/\sigma_P^2$ is .28 indicating that for large n and when the universe is normal, samples of size $.28n$ have the same degree of stability in setting tolerance ranges (10) as a sample of size n has when (x_r, x_{n-r+1}) is taken as the tolerance range. The same thing may be viewed in another way: The fact that the range of values of P' is 0 to 1 suggests that we may be able to get a fairly close approximation to the true distribution of P' by fitting a Pearson Type I function of the form

$$(13) \qquad \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, P'^{\alpha-1}(1-P')^{\beta-1},$$

determining α and β by equating the mean and variance of the distribution (13) to the mean and variance of P' respectively. Accordingly we find

$$(14) \qquad \begin{aligned} \alpha &= [a^2(1-a) - a\sigma_{P'}^2]/\sigma_{P'}^2 \\ \beta &= [a(1-a)^2 - (1-a)\sigma_{P'}^2]/\sigma_{P'}^2. \end{aligned}$$

Thus it will be seen from (14) that in order for the fitted distribution (13) to be identical with the distribution (1) a sample of only $\frac{t_a^2 e^{-t_a^2}}{\pi a(1-a)}(n+2)$ cases is needed.

In case only one tolerance limit is to be set, e.g. $\bar{x} - t_a\sqrt{(n+1)/n}\cdot s$, the

proportion, say u', of the universe which will be included in the tail has mean value $(1 - a)/2$ and variance $\frac{(2 + t_a^2)}{4\pi n}$ approximately for large n. The ratio of this variance to that of u, which is approximately $(1 - a^2)/4n$ for large n, gives the efficiency of using x_r for the lower tolerance limit in case of a normal universe. For example, if $a = .99$, the efficiency is .18.

It is perhaps appropriate here to point out the distinction between confidence limits and tolerance limits. It is well-known that in a sample from a normal universe with mean m the probability is a that the confidence limits $\bar{x} \pm t_a s$ will include the population mean m between them. The tolerance limits $\bar{x} \pm t_a\sqrt{(n+1)/n}\cdot s$, on the other hand are used to estimate the middle $100a\%$ of the universe. Although the tolerance limits $\bar{x} \pm t_a\sqrt{(n+1)/n}\cdot s$ are much more stable for a given sample size than those given by x_r and x_{n-r+1}, in case of a normal distribution, it should be emphasized that in case of even slight non-normality, particularly when skewness is present, the former pair of limits are apt to give very erroneous results with reference to the proportion of the universe included in the tails. Confidence limits estimating m are probably much less sensitive to skewness than tolerance limits estimating the middle $100a\%$ of the universe, particularly when a is nearly unity.

Another important aspect of the problem of setting tolerance limits is the following: Suppose small samples of a given size are taken from a universe under statistical control. How many of these small samples should be taken as a basis for determining tolerance limits L_1 and L_2 of some function, say g, of the samples (e.g. the sum of the measurements in each sample) so that the proportion of samples in the universe of such samples having values of g between L_1 and L_2 will have a given mean with a given degree of stability? One obvious approach to this question is to consider a universe of samples in the same manner in which we have considered a universe of individuals throughout the present paper. This approach, however, does not make very efficient use of the observations, but we shall not enter into a treatment of the problem here. This problem and various related problems in the statistical methods of mass production remain to be studied.

4. Summary. A method based on truncated sample ranges for determining size of sample required for setting tolerance limits on a random variable x having any unknown continuous distribution $f(x)$ and having a given degree of stability is given. A method for setting tolerance limits corresponding to a given degree of stability in case $f(x)$ is normal is discussed and a comparison of the stabilities of the tolerance limits set by the two methods in the normal case is made. Illustrative examples of the methods are given.

ON A CERTAIN CLASS OF ORTHOGONAL POLYNOMIALS

By Frank S. Beale

Lehigh University, Bethlehem, Pennsylvania

Introduction. E. H. Hildebrandt has demonstrated the following theorem[1]: *If y is a non-identically zero solution of the Pearsonian Differential Equation,*

$$\frac{1}{y}\frac{dy}{dx} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2} \equiv \frac{N}{D}, \qquad a_i, b_i \text{ real, then} \tag{1}$$

$$\frac{D^{n-k}}{y}\frac{d^n}{dx^n}(D^k y) \equiv P_n(k, x), \qquad n, k \text{ integers, } n \geq 0, \text{ is a} \tag{2}$$

polynomial in x of degree n at most. Hildebrandt has obtained various relations connecting the $P_n(k, x)$ and their derivatives as well as a recurrence relation.

If in (2) we set $k = n$ there results from a proper choice of N and D in (1), the classical Hermite, Laguerre, Jacobi and Legendre Polynomials. Many properties of these classical polynomials have been obtained by numerous investigators.[2]

One of the most important of these properties is that of orthogonality which can be stated as follows: *Consider a sequence of the classical polynomials $\Phi_i(x) = x^i - S_i x^{i-1} + \cdots$. There exists an interval (a, b) finite or infinite and a unique weight function $\psi(x)$, monotonic non-decreasing over (a, b) such that,*

$$\int_a^b \Phi_m(x)\Phi_n(x)\, d\psi(x) = 0, \qquad \text{for } n \neq m. \tag{3}$$

In the future we will refer to the type of orthogonality given by (3) *with $\psi(x)$ monotonic non-decreasing as orthogonality in the restricted sense.* In order to determine whether a given system of polynomials is orthogonal in the restricted sense we have the following theorem:[3]

Theorem 1. *In order that the sequence of polynomials $\Phi_i(x) = x^i - S_i x^{i-1} +$*

---

[1] E. H. Hildebrandt, "Systems of polynomials connected with the Charlier expansions, etc.," *Annals of Math. Stat.*, Vol. 2(1931), pp. 379–439.

[2] For an account of these properties as well as an extensive bibliography the reader can refer to one of two treatises viz : J. Shohat, *Théorie Générale des Polynomes Orthogonaux de Tchebichef*, Memoriale des Sciences Mathématiques, Fascicule 66, Paris, Gauthier Villars, 1936.

Gabor Szegö, *Orthogonal Polynomials*, Am. Math. Soc., Colloquium Publications, Vol. 23, 1939.

[3] J. Shohat, "The relation of the classical orthogonal polynomials to the polynomials of Appell," *Am. Jour. of Math.*, Vol. 58(1936), pp 454–455.

$\cdots$, $i = 1, 2, 3, \cdots$ *with real coefficients be orthogonal in the restricted sense it is necessary and sufficient that there exist a recurrence relation,*

$$(4) \qquad \Phi_i(x) = (x - c_i)\Phi_{i-1}(x) - \lambda_i\Phi_{i-2}(x), \qquad \Phi_0 = 1, \qquad \Phi_1 = x - c_1,$$

c_i, λ_i *const. with all* $\lambda_i > 0$, $i \geq 2$.

With Shohat[4] *we will say that a system of polynomials* $\Phi_i(x) = x^i - S_1x^{i-1} + \cdots$, $i = 1, 2, 3, \cdots$, *with real coefficients is orthogonal in the general sense if there exists at least one weight function* $\psi(x)$, *of bounded variation over* (a, b) *such that* (3) *is satisfied.* In connection with generalized orthogonality we have the following theorem:[4]

THEOREM 2. *In order that the system* $\Phi_i(x)$, $i = 1, 2, 3, \cdots$ *be orthogonal in the general sense it is necessary and sufficient that relation* (4) *be satisfied with all* $\lambda_i \neq 0$.

It is the purpose of this paper to investigate the orthogonality properties of the general polynomials $P_n(n, x)$ given by (2). In Part 1 a general recurrence relation is derived which applies to all the polynomials $P_n(k, x)$. In Part 2 all the different types of orthogonal polynomials $P_n(n, x)$ are determined by making use of the general recurrence relation derived in Part 1. We also show, following lines laid down by Hahn[5], that the only systems of polynomials with simple zeros which are orthogonal in either the restricted or the general sense and whose derivatives are orthogonal in either sense are the systems considered in Part 2.

1. **The general recurrence relation.** From (2) we can write,

$$(5) \qquad P_{n-1}(k, x) = \frac{D^{n-k-1}}{y}\frac{d^{n-1}}{dx^{n-1}}D^k y = \frac{D^{n-k-1}}{y}\frac{d^{n-1}}{dx^{n-1}}[D \cdot D^{k-1}y].$$

Apply Leibnitz Formula to the right side and make use of (2). There results,

$$(6) \qquad \begin{aligned} P_{n-1}(k, x) = {} & P_{n-1}(k-1, x) + (n-1)D'P_{n-2}(k-1, x) \\ & + \frac{(n-1)(n-2)}{1\cdot 2}D''DP_{n-3}(k-1, x). \end{aligned}$$

From Hildebrandt's paper we have,[6]

$$(7) \qquad P_{n+1}(k+1, x) = [N + (k+1)D']P_n(k, x) + n[N' + (k+1)D'']DP_{n-1}(k, x).$$

Decrease k and n each by one in (7) and obtain a relationship which we number (8). Again decrease n by one in (8) and get a relation which we number (9).

---

[4] J. Shohat, "Sur les polynomes orthogonaux généralisés," *Comptes Rendus*, Vol. 207 (1938), p. 556.

[5] Wolfgang Hahn, "Über die Jacobischen polynome und zwei verwandte polynomklassen," *Math. Zeits.*, Vol. 39(1934-35), pp. 634-638.

[6] E. H. Hildebrandt, loc cit p. 407.

From (6), (7), (8) and (9) eliminate $P_{n-1}(k, x)$, $P_{n-2}(k-1, x)$, and $P_{n-2}(k-1, x)$. There results,

$$
\begin{aligned}
(10)\qquad & [2N' + (2k - n + 1)D''][N' + kD'']P_{n+1}(k+1, x) \\
&= \{[2N' + (2k - n + 1)D''][N' + kD''][N + (k+1)D'] \\
&\quad + n[N' + (k+1)D''][2N'D' + kD'D'' - ND'']\}P_n(k, x) \\
&\quad + n[N' + (k+1)D'']\{2(N' + kD'')^2 D \\
&\quad - (N + kD')(2N'D' + kD'D'' - ND'')\}P_{n-1}(k-1, x).
\end{aligned}
$$

In (10) decrease n and k each by one and replace N and D by their values from (1). Thus we get,

$$
\begin{aligned}
(11)\qquad & [a_1 + (2k - n)b_2][a_1 + 2(k-1)b_2]P_n(k, x) \\
&= \{[a_1 + (2k-2)b_2][a_1 + 2kb_2][a_1 + (2k-1)b_2]x \\
&\quad + [a_1 + (2k-2)b_2][a_1 + (2k-n)b_2][a_0 + kb_1] \\
&\quad + (n-1)[a_1 + 2kb_2][a_1b_1 + (k-1)b_1b_2 - a_0b_2]\}P_{n-1}(k-1, x) \\
&\quad + (n-1)[a_1 + 2kb_2]\{b_0[a_1 + (2k-2)b_2]^2 \\
&\quad - [a_0 + (k-1)b_1][a_1b_1 + (k-1)b_1b_2 - a_0b_2]\}P_{n-2}(k-2, x).
\end{aligned}
$$

In this recurrence formula the $P_n(k, x)$ have in general a coefficient of x^n different from one. *Polynomials which have one for the coefficient of x^n we will refer to in the future as normalized.* Let us now transform (11) for normalized $P_n(k, x)$. Theorem 1 deals with polynomials normalized in the above sense. Let us write,

$$P_n(k, x) = a_{n,k}x^n - b_nx^{n-1} + \cdots. \quad \text{In (4) set,} \quad \Phi_n(x) = P_n(k, x)/a_{n,k}.$$

Thus we get,

$$(12)\qquad P_n(k, x) = (A_nx - B_n)P_{n-1}(k-1, x) - \gamma_nP_{n-2}(k-2, x)$$

where

$$\gamma_n \equiv \frac{a_{n,k}}{a_{n-2,k-2}}\lambda_n, \qquad A_n \equiv \frac{a_{n,k}}{a_{n-1,k-1}}, \quad \text{and} \quad B_n \equiv \frac{a_{n,k}}{a_{n-1,k-1}}C_n.$$

Relation (12) is essentially of the same form as (11). Each of these is to be reduced to form (4).

From a previous paper by the author[7] we have,

$$(13)\qquad P'_{n+1}(k, x) = (n+1)[N' + \tfrac{1}{2}(2k - n)D'']P_n(k, x).$$

$n - 1$ successive applications of this relation give us, $[P_0(k, x) \equiv 1]$, that the coefficient of x^n in $P_n(k, x)$ is,

[7] Frank S. Beale, "On the polynomials related to Pearson's differential equation," *Annals of Math. Stat.*, Vol. 8(1937), p. 207 (2).

$$a_{n,k} = \prod_{i=0}^{n-1} [a_1 + (2k - n + 1 + i)b_2]. \tag{14}$$

By employing (14) in (12) we see that (12) or (11) reduces to form (4) where,

$$c_n = -\frac{[a_1 + (2k - n)b_2][a_0 + kb_1]}{[a_1 + 2kb_2][a_1 + (2k - 1)b_2]} - (n - 1)\frac{[a_1b_1 + (k - 1)b_1b_2 - a_0b_2]}{[a_1 + (2k - 1)b_2][a_1 + (2k - 2)b_2]}. \tag{15}$$

$$\lambda_n = -(n - 1)\frac{[a_1 + (2k - n - 1)b_2]\{b_0[a_1 + (2k - 2)b_2]^2 - [a_0 + (k - 1)b_1][a_1b_1 + (k - 1)b_1b_2 - a_0b_2]\}}{[a_1 + (2k - 3)b_2][a_1 + (2k - 2)b_2]^2[a_1 + (2k - 1)b_2]} \tag{16}$$

Equation (16) together with Theorems 1 and 2 can now be applied to the polynomials $P_n(k, x)$.

From (14) *it is seen that* $P_n(k, x)$ *is of degree* n *provided that none of the factors of the product vanishes. This condition we assume to hold here for all* n.

We can now obtain a recurrence relation for the qth derivatives of $P_n(k, x)$. A repeated application of (13) leads to,

$$\frac{d^q}{dx^q} P_n(k, x) = P_{n-q}(k, x) \prod_{i=0}^{q-1} (n - i)\,[a_1 + (2k - n + i + 1)b_2], \tag{17}$$

where $P_n(k, x)$ is not normalized in the above sense. By considering the right side of (17) together with (14) we see that (17) can be divided by

$$a_{n-q,k} \prod_{i=0}^{q-1} (n - i)\,[a_1 + (2k - n + i + 1)b_2]$$

and thus normalize the polynomials on both the right and left sides of (17). Consequently the recurrence relation for normalized $d^q[P_n(k, x)]/dx^q$, $n = 0, 1, 2, \cdots$, is identical with the recurrence relation for normalized $P_{n-q}(k, x)$ as given by (4), (15) and (16) when we replace n by $n - q$ in these latter.

2. The different types of orthogonal $P_n(n, x)$. Suppose first that $b_2 \neq 0$ in (1). A transformation on x with real coefficients can be affected which changes (1) into either,

$$\frac{1}{y}\frac{dy}{dx} = \frac{(\alpha - \beta) + (-\alpha - \beta)x}{1 - x^2} \quad \text{or} \tag{18}$$

$$\frac{1}{y}\frac{dy}{dx} = \frac{-2mx - q}{a^2 + x^2}. \tag{19}$$

(A) Equation (18) together with (2) for $k = n$ defines the generalized Jacobi Polynomials (normalized in the above sense),

$$J_n(x, \alpha, \beta) = \frac{1}{a_{n,n}}(1 + x)^{-\alpha}(1 - x)^{-\beta}\frac{d^n}{dx^n}[(1 + x)^{n+\alpha}(1 - x)^{n+\beta}]$$

where $1/a_{n,n}$ is given by (11). If in (16) we set $k = n$ and make proper replacements for constants as (18) and (1) show we have,

$$\lambda_n = 4(n-1)\frac{(\alpha+\beta+n-1)(\alpha+n-1)(\beta+n-1)}{(\alpha+\beta+2n-3)(\alpha+\beta+2n-2)^2(\alpha+\beta+2n-1)}, \tag{20}$$

$$n \geq 2.$$

From Theorem 1 and this value of λ_n we conclude that if $\alpha > -1$, $\beta > -1$, the sequence $\{J_n(x, \alpha, \beta)\}$ is orthogonal in the restricted sense—a well-known result. From Theorem 2 we can similarly conclude that if neither α, β, nor $(\alpha + \beta)$ equals $-j$, j a positive integer, the sequence $\{J_n(x, \alpha, \beta)\}$ is orthogonal in the general sense.

(A_1) If in (18) we set $\alpha = \beta = 0$ we obtain a differential equation which together with (2) for $k = n$ leads to the Legendre Polynomials, (normalized in above sense), $P_n(x) = \frac{n!}{(2n)!}\frac{d^n}{dx^n}(x^2-1)^n$. Setting $\alpha = \beta = 0$ in (20) leads to $\lambda_n = \frac{(n-1)^2}{(2n-3)(2n-1)}$, $n \geq 2$. Thus from Theorem 1 we conclude that the Legendre Polynomials are orthogonal in the restricted sense, a result well known.

(B) Equation (19) together with (2) for $k = n$ leads to a class of polynomials (normalized in above sense), mentioned by Romanovsky.[8]

$$R_n(x, m, q, a) = \frac{1}{a_{n,n}}(a^2+x^2)^m \exp\left(\frac{q}{a}\tan^{-1}\frac{x}{a}\right)\frac{d^n}{dx^n}\left[(a^2+x^2)^{n-m}\exp -\frac{q}{a}\tan^{-1}\frac{x}{a}\right]$$

where again $1/a_{n,n}$ is given by (14). In (16) set $k = n$ and make the proper replacements of constants and,

$$\lambda_n = \frac{n-1}{4}\frac{(2m-n+1)\{4a^2(m-n+1)^2+q^2\}}{(2m-2n+3)(m-n+1)^2(2m-2n+1)}, \qquad n \geq 2.$$

From Theorem 2 it now follows that the sequence $\{R_n(x, m, q, a)\}$ is orthogonal in the general sense if $m \neq j/2$, j a positive integer. There is no set of parameters m, q, a which assures orthogonality in the restricted sense.

In connection with Romanovsky's note there appear to be several discrepancies. For the weight functions given there under types IV and V, the nth moments for sufficiently large n do not exist over the intervals there considered. Type V is the special case of type IV for $a = 0$. Type VI is none other than Jacobi Polynomials so that the orthogonality relations given there for this case are incorrect. In all three types listed certain of the recurrence relations for the polynomials are in error.

(B_1) We note here one special sub-class of R_n. Take $m = q = 0$ and $a = 1$ in (19). We obtain from (2) and (14) a system of normalized polynomials analogous to the Legendre Polynomials namely, $\phi_n(x) = \frac{n!}{(2n)!}\frac{d^n}{dx^n}(x^2+1)^n$.

[8] V. Romanovsky, "Sur quelques classes nouvelles de polynomes orthogonaux," *Comptes Rendus*, Vol. 188(1929), pp. 1023–1025.

It is easy to verify for these that,

$$\int_{-i}^{i} \phi_n(x)\phi_m(x)\,dx = 0, \qquad m \neq n, \qquad i = \sqrt{-1}.$$

(C) Suppose that in (1), $b_2 = 0$, $b_1 \neq 0$. A linear transformation with real coefficients changes (1) into, $\frac{1}{y}\frac{dy}{dx} = \frac{\alpha - x}{x}$. This equation together with (2) and (14) for $k = n$ defines the generalized Laguerre Polynomials, (normalized in above sense), $L_n(x, \alpha) = (-1)^n x^{-\alpha} e^{x} \frac{d^n}{dx^n}[x^{n+\alpha} e^{-x}]$. Setting $k = n$ and making proper replacements in (16) we get, $\lambda_n = (n-1)(\alpha + n - 1)$, $n \geq 2$. From Theorem 1 we see that if $\alpha > -1$ the L_n are orthogonal in the restricted sense, a well-known result. From Theorem 2 we can say that if $\alpha \neq -j$, j a positive integer, the polynomials are orthogonal in the general sense.

(D) If in (1), $b_1 = b_2 = 0$, $b_0 \neq 0$ we can perform a linear transformation on x with real coefficients and get, $\frac{1}{y}\frac{dy}{dx} = hx$. This differential equation together with (2) and (14) gives a set of normalized polynomials $G_n(x) = \frac{1}{h^n} e^{hx^2/2} \frac{d^n}{dx^n} e^{-hx^2/2}$. Taking $k = n$ and making proper substitutions for constants in (16) we get $\lambda_n = -(n-1)/h$, $n \geq 2$. If h is negative it follows from Theorem I that the sequence $\{G_n(x)\}$ is orthogonal in the restricted sense. In fact, $G_n(x) = H_n(x)$ = Hermite Polynomials.

On the other hand, if h is positive we have from Theorem 2 orthogonality in the general sense. In fact, it can be easily verified for this case that,

$$\int_{-i\infty}^{i\infty} e^{hx^2/2} G_n(x)G_m(x)\,dx = 0, \qquad m \neq n, \qquad i = \sqrt{-1}.$$

(E) The only remaining possibility for (1) not so far discussed occurs when $N \equiv$ constant and D is linear. In this case it has been shown that $P_n(k, x)$ of (2) reduces to a constant.[9]

E. H. Hildebrandt has shown[10] that the polynomials $P_n(n, x)$ of (2) satisfy a differential equation of the form,

$$(21) \qquad (b_0 + b_1 x + b_2 x^2)\frac{d^2y}{dx^2} + [a_0 + b_1 + (a_1 + 2b_2)x]\frac{dy}{dx} - n[a_1 + (n+1)b_2]y = 0, \qquad n = 1, 2, 3, \ldots.$$

Moreover with the coefficients of d^2y/dx^2 and dy/dx in (21) he has shown that for (21) to have a polynomial solution of degree n the coefficient of y must be of the form given in (21).

[9] Frank S. Beale, loc. cit. p. 209, Theorem I_9.

[10] Loc. cit. pp. 404–405.

From (16) we can say that for $k = n$ and an orthogonal sequence $P_n(n, x)$, $n = 0, 1, 2, \cdots$ we have,

$$a_1 + (n-1)b_2 \neq 0, \tag{22}$$

$$b_0[a_1 + (2n-2)b_2]^2 \cdot [a_0 + (n-1)b_1][a_1b_1 + (n-1)b_1b_2 - a_0b_2] \neq 0, \tag{23}$$

where n is an integer ≥ 2. Considering for (21) a solution of the type $y = \sum_{i=0}^{\infty} c_i x^i$ we readily show that if (22) and (23) are satisfied, (21) possesses for each n a single polynomial solution of degree n. Two solutions which differ merely by a constant factor are regarded as the same solution. This polynomial solution of (21) must be $P_n(n, x)$.

By employing theorems from a previous paper by the author[11] we can show that if (22) and (23) are satisfied, the zeros of the polynomials of section II are simple whether these zeros are real or complex.

Hahn has shown[12] that if a set of normalized polynomials and their derivatives satisfy a relation of the form (4) with $\lambda_i \neq 0$ and if the zeros of the polynomials are all simple then the polynomials must necessarily satisfy an equation of form (21). Since in this paper we have considered all possible values of a_i, $(i = 0, 1)$, and b_i, $(i = 0, 1, 2)$, which lead to orthogonal polynomials, it follows that the only systems of polynomials with simple zeros and orthogonal in either restricted or general sense whose derivatives in turn are orthogonal in either sense are the systems of section 2.

[11] Loc. cit. pp. 207-209, Theorems I_1 to I_{10}.

[12] Loc. cit. pp. 634-636.

THE SKEWNESS OF THE RESIDUALS IN LINEAR REGRESSION THEORY

By P. S. Dwyer

University of Michigan, Ann Arbor, Mich.

In obtaining the regression of y on x it is customary to show the relation between the actual and the estimated y by computing the standard deviation of the residuals with the use of the formula $\sigma_\epsilon = \sigma_y \sqrt{1 - r^2}$. If the errors are distributed normally one may estimate the number of values coming within one standard deviation, within two standard deviations, etc., of the regression line. However these errors are not always distributed normally, and in such a case it seems wiser to compute the skewness of the residuals and to use a Pearson Type III curve in making the interpretation. The present paper outlines a technique for the calculation of $\alpha_{3:\epsilon}$ which is feasible from a practical standpoint. It is based (*a*) on a cumulative totals method of obtaining the correlation coefficient which, at the same time, makes possible the determination of the third order moments needed to evaluate the skewness and (*b*) on an efficient ritual for computing the coefficient of skewness from the moments.

The determination of the normality or non-normality of the residuals is not always immediately evident. If the scatter diagram or correlation chart is presented, one can make an estimate of the extent of normality but if not, and the most modern and efficient computational methods do not utilize the correlation chart, there is no way by which the presence or absence of normality can be detected. Some research workers are opposed to the use of the more efficient methods (particularly the use of the Hollerith tabulators) because the correlation chart is not presented. Though within limits it is possible to use the tabulator to present the correlation chart simultaneously with the values needed to compute the correlation coefficient [1], it is here suggested that the computation of the skewness of residuals, which can now be accomplished quite easily from the tabulator runs, may be substituted for the examination of the correlation chart.

The classical least squares theory makes use of

$$\epsilon = y - b_0 - b_1 x \tag{1}$$

where b_0 and b_1 are the solutions of the normal equations. We note that the first normal equation is $\Sigma\epsilon = 0$ so that $M_\epsilon = 0$ and the residual is a deviation. It follows that the skewness of residuals is

$$\alpha_{3:\epsilon} = \frac{\Sigma(y - b_0 - b_1 x)^3}{N\sigma_\epsilon^3}. \tag{2}$$

We wish to compute $\alpha_{3,\epsilon}$ without computing the individual residuals. The denominator causes us little concern but it seems discouraging to evaluate such an expression as

$$\Sigma y^3 - Nb_0^3 - b_1^3\Sigma x^3 - 3b_0\Sigma y^2 - 3b_1\Sigma xy^2 + 3b_0^2\Sigma y - 3b_0^2b_1\Sigma x + 3b_1^2\Sigma x^2 y - 3b_1^2b_0\Sigma x^2 + 6b_0b_1\Sigma xy$$

even though the values of b_0, b_1, N, Σx, Σy, Σx^2, Σxy, Σy^2, Σx^3, $\Sigma x^2 y$, Σxy^2, Σy^3 are available.

A first simplification is made by summing (1) and dividing by N. We then have

$$M_\epsilon = M_y - b_0 - b_1 M_x \tag{3}$$

and by subtracting (3) from (1) and denoting deviations by barred letters, we have

$$\epsilon = \bar{y} - b_1\bar{x} \tag{4}$$

so that the skewness of errors is

$$\alpha_{3,\epsilon} = \frac{\Sigma\bar{y}^3 - 3b_1\Sigma\bar{x}\bar{y}^2 + 3b_1^2\Sigma\bar{x}^2\bar{y} - b_1^3\Sigma\bar{x}^3}{N\sigma_\epsilon^3}. \tag{5}$$

This formula can also be expressed as

$$\alpha_{3:\epsilon} = \frac{\bar{\mu}_{03} - 3b_1\bar{\mu}_{12} + 3b_1^2\bar{\mu}_{21} - b_1^3\bar{\mu}_{30}}{[\bar{\mu}_{02} - b_1\bar{\mu}_{11}]^{3/2}}. \tag{6}$$

A similar formula for the skewness of the residuals of x on y is

$$\alpha_{3:\epsilon'} = \frac{\bar{\mu}_{30} - 3b_1'\bar{\mu}_{21} + 3b_1'^2\bar{\mu}_{12} - b_1'^3\bar{\mu}_{03}}{[\bar{\mu}_{20} - b_1'\bar{\mu}_{11}]^{3/2}}. \tag{7}$$

For theoretical purposes formula (6) may be put in standard units with $b_1 = r\frac{\sigma_y}{\sigma_x}$, $b_1' = r\frac{\sigma_x}{\sigma_y}$, $\bar{\mu}_{30} = \alpha_{30}\sigma_x^3$, $\bar{\mu}_{21} = \alpha_{21}\sigma_x^2\sigma_y$, etc. with the resulting

$$\alpha_{3:\epsilon} = \frac{\alpha_{03} - 3r\alpha_{12} + 3r^2\alpha_{21} - r^3\alpha_{30}}{(1 - r^2)^{3/2}}. \tag{8}$$

As $r \to 0$, $\alpha_{3,\epsilon} \to \alpha_{3,y}$ just as $\sigma_\epsilon \to \sigma_y$ as $r \to 0$.

Formulas (6) and (7) are of some theoretical importance in that they show how the skewness of the residuals is connected with the skewness of the marginal distribution. Thus

as $\bar{\mu}_{11} \to 0$, b_1 and $b_1' \to 0$ and $\alpha_{2,\epsilon} \to \alpha_{3,y}$, $\alpha_{3:\epsilon'} \to \alpha_{3,x}$;

as $b_1 \to \infty$, $\alpha_{3:\epsilon} \to -\alpha_{3:x}$ and as $b_1' \to \infty$, $\alpha_{3:\epsilon'} \to -\alpha_{3,y}$;

as $b_1 \to 1$, $\alpha_{3:\epsilon} \to \alpha_{3:y-x}$. Similarly as $b_1' \to 1$, $\alpha_{3:\epsilon'} \to \alpha_{3,x-y}$.

It is hence possible in some cases to get a good approximation to the skewness of the residuals if the regression coefficients and the skewness of the marginal distribution are known.

TABLE I

Correlation from first order cumulations

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Y \ X			4.00	3.99 3.50-	3.49 3.00-	2.99 2.50-	2.49 2.00-	1.99 1.50-	1.49 1.00	.99 .50	.49 .00		
	y \ x		8	7	6	5	4	3	2	1	0		
		f_x \ f_y	13	50	107	220	341	179	121	60	35	Cx_y	Cy_y
4.00	6	18	5	2	5	5	1					113	108
3.99 3.50-	5	106	2	19	29	27	20	7		1	1	673	638
3.49 3.00-	4	178	3	12	35	53	44	18	6	5	2	1503	1350
2.99 2.50-	3	270	3	10	20	55	103	33	27	11	8	2568	2160
2.49 2.00-	2	330		6	11	54	114	67	46	19	13	3714	2820
1.99 1.50-	1	173		1	5	19	45	44	34	18	7	4244	2993
1.49 1 00-	0	51			2	7	14	10	8	6	4	4309	2993
		Cy_x	61	259	661	1330	2194	2578	2809	2923	2993	12815	10069
		Cx_x	104	454	1096	2196	3560	4097	4339	4399	4399	20245	

For actual computation, we use (6) and (7). It has been indicated previously how the values Σx, Σy, Σx^2, Σxy, Σy^2, Σx^3 and Σy^3 could be obtained with the use of cumulations. An illustration used previously [2] is presented in Table I. The information was obtained from the Office of Educational Investigations of the University of Michigan and gives the University first semester average (X) and the high school average (Y) for 1,126 students entering the College of Literature, Science, and the Arts in 1928.

The new origin of each variable is taken at the class mark of the lowest class rather than at the class mark of a middle class as is conventional. In this way all negative terms are avoided in the computation of the moments. The x's are arranged in descending order from left to right and the y's in descending order from top to bottom. The notation x_y is used to indicate the sum of all the x's

having the same value of y. Thus the first entry in column 13 is $5\cdot8 + 2\cdot7 + 5\cdot6 + 5\cdot5 + 1\cdot1 = 113$. The column Cx_y is obtained by cumulating the values of x_y. Similarly y_y is the sum of all the y's having the same value and the first entry in column 14 is $18(6) = 108$. The entries Cy_y, Cy_x, and Cx_x are obtained similarly.

The entries Σx, Σy, Σx^2, Σxy, Σy^2 are found in the lower right hand box in this position:

	Σx	Σy
Σy	Σxy	Σy^2
Σx	Σx^2	

The values of Σx and Σy are obtained from the final cumulations while the value of Σxy is obtained by adding the entries in the column above, or, as a check, the entries in the row to the left. The value of Σy^2 is obtained by adding the entries in the row at the left of the box while the value Σx^2 is obtained by adding the entries above the box.

The values of the third order sums are obtained by multiplying the entries above the box and to the left of the box successively by 1, 3, 5, 7, 9, etc. Thus,

$$
\begin{aligned}
\Sigma x^3 &= 4399 + 3(4339) + 5(4097) + \text{etc.} = 102{,}103,\\
\Sigma x^2 y &= 2923 + 3(2809) + 5(2578) + \text{etc.} = 63{,}121,\\
\Sigma x y^2 &= 4244 + 3(3714) + 5(2568) + \text{etc.} = 46{,}047,\\
\Sigma y^3 &= 2993 + 3(2820) + 5(2160) + \text{etc.} = 38{,}633.
\end{aligned} \tag{9}
$$

In making the reductions we use $ab - cd$ operations as much as possible. We first compute

$$
\begin{aligned}
A_{x,y} &= N\Sigma xy - (\Sigma x)(\Sigma y),\\
A_{x,x} &= N\Sigma x^2 - (\Sigma x)^2,\\
A_{x^2,y} &= N\Sigma x^2 y - (\Sigma x^2)(\Sigma y).
\end{aligned} \tag{10}
$$

We note too that

$$
\begin{aligned}
\mu_{30} &= [NA_{x^2,x} - (2\Sigma x)(A_{x,x})]/N^3; \qquad \mu_{21} = [NA_{x^2,y} - (2\Sigma x)(A_{x,y})]/N^3\\
\mu_{12} &= [NA_{x,y^2} - (2\Sigma y)(A_{x,y})]/N^3; \qquad \mu_{03} = [NA_{y^2,y} - (2\Sigma y)(A_{y,y})]/N^3
\end{aligned} \tag{11}
$$

and finally we get $\alpha_{3:4}$ or $\alpha_{3:4'}$ by (6) or (7).

The general solution is outlined on the left of Table II. We record in Fig. A the values given by (9) and in the Fig. B the values resulting from the application of (10). The values $2\Sigma y$ and $2\Sigma x$ are inserted in Fig. B to facilitate the calculation of Fig. C which gives the values of (11). The technique is very easily carried out once it is understood. It can be performed with hand calcu-

lators but it is ideally adapted to the use of the latest Marchant, Fridén, and Monroe models equipped with automatic positive and negative multiplication, so that $ab \cdot cd$ operations can be performed with a minimum of effort and a maximum of accuracy. Actually the value of "a," which is the total frequency, is the same for many of these operations so that there is further saving if a machine is used which permits the locking in of a constant in such a way that it can be used, without continued key punching, in later $ab \cdot cd$ operations.

TABLE II

Abbreviated techniques for computing third order central moments, etc.

Fig. A.

N	Σx	Σx^2	Σx^3	1126	4790	20215	102103
Σy	Σxy	Σx^2y		2893	12815	63121	
Σy^2	Σxy^2			10069	46017		
Σy^3				38633			

Fig. B.

N	$2\Sigma x$	$A_{x,x}$	$A_{x^2,x}$	1126	8798	3414069	25910223
$2\Sigma y$	$A_{x,y}$	$A_{x^2,y}$		5986	1263483	10180961	
$A_{y,y}$	$A_{y^2,x}$			2379645	7555391		
$A_{y^2,y}$				13304211			

Fig. C.

N		$A_{x,x}$	$N^3\mu_{30}$	1126		3414069	1131286704
	$A_{x,y}$	$N^3\mu_{12}$			1263483	685438652	
$A_{y,y}$	$N^3\mu_{21}$			2379645	944161028		
$N^3\mu_{03}$				803580396			

Fig. D.

N	(b_1)	μ_{20}	μ_{03}	1126	(.367)	2.717	−.7925
(b_1')	μ_{11}	μ_{21}	$(-b_1^2)$, $(-3b_1')$	(531)	.997	.1901	(−1 593)
μ_{02}	μ_{12}	$(3b_1^2)$, $(3b_1'^2)$		1.877	.6614	(846)	
μ_{03}	$(-3b_1)$, $(-b_1'^3)$			.5620	(−.150)		

The values in Fig. D are obtained by dividing the values $A_{y,y}$, $A_{x,y}$, and $A_{x,x}$ in Fig. C by N^2 and the values in the diagonal below, $NA_{y^2,y} - (2\Sigma y)A_{y,y}$, etc., by N^3. The values $b_1 = \frac{\bar{\mu}_{11}}{\bar{\mu}_{20}}$ and $b_1' = \frac{\bar{\mu}_{11}}{\bar{\mu}_{02}}$ can be inserted in Fig. D adjacent to the N. The value of the correlation coefficient is $r = \sqrt{b_1 b_1'} = \frac{\bar{\mu}_{11}}{\sqrt{\bar{\mu}_{20}\bar{\mu}_{02}}}$.

We have $\sigma_x = \sqrt{\mu_{x2}} = b_1\bar{\mu}_{11}$ and $\sigma_{x'} = \sqrt{\bar{\mu}_{x2}} = b_1'\bar{\mu}_{11}$ so that the standard deviation of residuals is readily computed from the entries of Fig. D. The numerator of (6) is readily obtained after entering $-3b_1$, $3b_1^2$, $(-b_1^3)$ in the diagonal under the diagonal containing the third moments and multiplying by columns. The numerator of (7) is obtained by entering $-b_1'^2$, $3b_1'^2$, $-3b_1'$, in the same diagonal and multiplying by rows. The theory is applied to the results of Table I and the details are presented at the right of Table II. It is to be noted that all values indicated here are the coded values x, y and not the original values, X, Y. However, the correlation coefficient and the skewness of errors are independent of any such change in unit, grouping errors being neglected.

From Fig. D we see that $b_1 = .997/2.717 = .367$, that $b_1' = .997/1.877 = .531$ and that $r = \sqrt{(.367)(.531)} = .441$. In this case we wish to estimate college record, x, from high school record, y, so we use $b_1' = .531$ and compute $-3b_1' = -1.593$, $3b_1'^2 = .846$, $-b_1'^3 = -.150$. It follows that

$$\alpha_{3,x'} = \frac{-.7925 + (.4801)(-1.593) + (.6614)(.846) + (.5629)(-.150)}{[2.717 - .531(.997)]^{3/2}} = -.334.$$

It thus appears that a better picture of the variation of the residuals in this case is obtained with the use of a Pearson Type III with α_3 approximately $-\frac{1}{3}$ than is obtained with the use of a normal curve. It is not necessary, of course, to form Fig. D as the results can all be obtained from Fig. C. Thus if we multiply the numerator and denominator of (6) by N^3, we get entries, with the exception of the b's, which are in Fig. C. Now in this case $b_1 = \frac{A_{x,y}}{A_{x,x}}$ and $b_1' = \frac{A_{x,y}}{A_{y,y}}$ so that these values can be inserted in the upper left as before. Also the powers of b_1' can be inserted in the lower right as in Fig. D. We have then

$$\alpha_{3,x'} = \frac{-1131,286,764 + (685,438,652)(-1.593) + (944,161,028)(.846) + (803,580,396)(-.150)}{[3444669 - (1263483)(.531)]^{3/2}}.$$

We know however, since the grades were coded, that it is not sensible to carry results to more than three places, (and, indeed, a three place determination of the skewness is very satisfactory for interpretive purposes even though more places might be obtained) so we cut down the number of places. The division of numerator and denominator by 10^6, and the dropping of the decimals results in

$$\alpha_{3,x'} = \frac{-1131 + 685(-1.593) + 944(.846) + 804(-.150)}{[344 - 126(.531)]^{3/2}} = -.335.$$

It is possible of course to duplicate the theory indicated in Table II with the use of moments rather than the A's. In this case Fig. A consists of 1, $\Sigma x/N$,

$\Sigma x^2/N$, etc. We have such formulas as $a_{x,y} = \dfrac{\Sigma xy}{N} - \dfrac{(\Sigma x)}{N}\dfrac{(\Sigma y)}{N} = \mu_{11} - \mu_{10}\mu_{01}$,

where $a_{xy} = \dfrac{A_{x,y}}{N_2}$, $a_{x^2,y} = \dfrac{A_{x^2,y}}{N_3}$, etc.

It would be possible to compute the $\alpha_{4,0}$ in a somewhat similar fashion though it would take somewhat longer. In the first place we would have to compute Σx^2y^2 from the correlation table. This could be done by forming the cumulation $C(y_x^2)$ and multiplying by 1, 3, 5, 7, 9, etc. When this is done, however, it does not appear that the calculation of the central moments of the fourth order can be reduced to as simple a ritual as the calculation of the third order moments.

The question should be raised as to the calculation of the skewness when there are two or more independent variables. This can be done, of course, but the calculations are lengthy. The point of the present paper is to provide an easy and simple technique for computing the skewness of residuals in the case of two variable linear regression.

REFERENCES

[1] Paul S. Dwyer and Alan D. Meacham, "The preparation of correlation tables on a tabulator equipped with digit selection," *Jour. Am. Stat. Ass.*, Vol 32 (1937), pp. 654-62. See particularly page 657.

[2] Paul S. Dwyer, "The computation of moments with the use of cumulative totals," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 288 304. See particularly pages 299 303.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

NOTE ON THE ADJUSTMENT OF OBSERVATIONS

By Arthur J. Kavanagh

The Forman Schools, Litchfield, Conn.

The method of least squares has been extended to the adjustment of observations with errors in more than one variable. The history of the development and its principal results have been given by Deming [2], [3], [4], [5]. The basis is the assumption that for the "best" adjustment the sum of the weighted squares of all the residuals (observed values minus adjusted values) must be made a minimum with respect to the adjustments to the observations and with respect to the parameters involved in the conditions the adjusted values must satisfy. In certain problems, such as some arising in the study of relative growth in biology, this assumption is not adequate; it is necessary that the sum to be minimized be generalized to include cross products as well as squares of the residuals.

Suppose we have a set of n universes of q-dimensional points whose centers of gravity are known to satisfy certain conditions; for instance, they might all lie on a certain type of curve. A sample having been taken from each universe, the center of gravity of each sample is taken as the observed center of gravity of the corresponding universe, and it is desired to determine the most probable set of adjustments to the coordinates and the most probable set of parameters involved in the conditions, subject to the requirement that the adjusted values satisfy the conditions exactly. It is assumed that the sampling distribution of the center of gravity in each universe satisfies the multivariate normal law, and that the standard deviations and coefficients of correlation of each sample may with sufficient accuracy be taken as the constants of the corresponding universe. Then by reasoning analogous to that of the derivation of the least squares principle for one variable from the univariate normal law, the probability of getting the observed set of values is proportional to e^{-Q}, where

$$Q = \sum_{i=1}^{n} Q_i \tag{1}$$

Q_i being a homogeneous quadratic function of the errors at the ith centroid and in general involving the cross products as well as the squares of the errors.

The probability will be a maximum when Q is a minimum. Consequently the best estimates for the coordinates of the centroids will be those making Q a minimum, subject to the conditions which the coordinates must satisfy.

For example it may be desired to study the relation between height and weight among growing boys by fitting a curve to the points whose abscissa and ordinate are respectively average height and average weight of a particular age group, one point corresponding to each age group in the study. The data for such a study are obtained from samples of the several age groups. Then the number n of universes is the number of age groups being studied, each universe consisting of the totality of two-dimensional points obtained by pairing the height with the weight of each boy in the age group. The centroid or "average point" of each universe would ideally be obtained from measurements of all the individuals of that age, but since sampling must be resorted to it is necessary to make allowances for the sampling distributions of the centroids. It is known that within each age group there is correlation between height and weight [1]. Consequently the sampling distribution of each centroid will exhibit a correlation which can be expressed in terms of the coefficient of correlation between height and weight of the individuals of the universe from which the sampling distribution arises. The existence of this correlation results in the presence of the cross-product term in the exponent of the bivariate normal formula describing the sampling distribution of the average values, that is in the Q_i of each centroid. If there were no such correlations the cross-product term in each Q_i would vanish and the situation would reduce to that of least squares.

In the general case, let $X_{1i}, X_{2i}, \cdots, X_{qi}$ be the observed coordinates of the ith centroid, $x_{1i}, x_{2i}, \cdots, x_{qi}$ the adjusted values (to be determined), and $V_{ji} = X_{ji} - x_{ji}$. Then Q_i may be written

$$
\begin{aligned}
Q_i = {} & w_{11i}V_{1i}^2 + w_{12i}V_{1i}V_{2i} + \cdots + w_{1qi}V_{1i}V_{qi} \\
& + w_{21i}V_{2i}V_{1i} + w_{22i}V_{2i}^2 + \cdots + w_{2qi}V_{2i}V_{qi} \\
& + \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
& + w_{q1i}V_{qi}V_{1i} + w_{q2i}V_{qi}V_{2i} + \cdots + w_{qqi}V_{qi}^2
\end{aligned} \tag{2}
$$

the w's being the weights, with $w_{jki} = w_{kji}$. Thus in the case of two variables, if N_i be the number of items in the ith sample, r_i its coefficient of correlation, and σ_{1i}, σ_{2i} its standard deviations, then

$$
w_{11i} = \frac{N_i}{2(1 - r_i^2)\sigma_{1i}^2}, \qquad w_{12i} = \frac{-N_i r_i}{2(1 - r_i^2)\sigma_{1i}\sigma_{2i}} = w_{21i}, \qquad w_{22i} = \frac{N_i}{2(1 - r_i^2)\sigma_{2i}^2}.
$$

The coefficients of the cross products in Q involve the coefficients of correlation of distributions. If the latter are all zero the cross products vanish and Q reduces to the sum of weighted squares, which is the basic expression of the least squares procedure. Consequently, from this point of view, the least squares assumption is equivalent to the assumption of zero correlation between the errors. The procedure in the more general situation might be called "least quadratics".

The Lagrange method of undetermined multipliers can be used to calculate the values of the adjustments to the coordinates and the values of the parameters. The procedure is the same as for least squares [2], [3], [5], the only difference being the somewhat greater complication of the algebra. We shall summarize the development here.

The condition equations, supposed ν in number, may be written

$$F^h(x_{i1}, \cdots, x_{in}; p_1, p_2, \cdots, p_r) = 0, \qquad h = 1, 2, \cdots, \nu,$$

where each F^h may in general involve any or all of the numbers x_{ji} as well as any or all of the parameters p_l, whose number we suppose to be r. Let

$$F^h_{ji} = \partial F^h / \partial x_{ji}, \qquad F^h_l = \partial F^h / \partial p_l \tag{3}$$

where the X's have been substituted for the x's after differentiation, and each p_l has been replaced by the best available approximate value p_{l0}. Let F^h_0 be the value of F^h after the same substitution. Also let $v_l = p_{l0} - p_l$. Then if the V's and v's are small the conditions may be written

$$\sum_i \sum_j F^h_{ji} V_{ji} + \sum_l F^h_l v_l = F^h_0, \qquad h = 1, 2, \cdots, \nu. \tag{4}$$

Differentiate Q with respect to the V's and equate the result to zero, eliminating the factor 2. Differentiate (4) with respect to the V's and the v's, multiply each equation by the corresponding undetermined multiplier $-\lambda_h$, and sum the results together with the result from differentiating Q. Collecting coefficients of the differentials δV_{ji} and δv_l, equating to zero and transposing the terms involving λ_h, we get

$$\begin{aligned} &w_{11i}V_{1i} + w_{12i}V_{2i} + \cdots + w_{1qi}V_{qi} = [\lambda_h F^h_{1i}] \\ &w_{21i}V_{1i} + w_{22i}V_{2i} + \cdots + w_{2qi}V_{qi} = [\lambda_h F^h_{2i}] \qquad (i = 1, 2, \cdots, n) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &w_{q1i}V_{1i} + w_{q2i}V_{2i} + \cdots + w_{qqi}V_{qi} = [\lambda_h F^h_{qi}] \end{aligned} \tag{5}$$

$$[\lambda_h F^h_l] = 0 \qquad l = 1, 2, \cdots, r, \tag{6}$$

where the brackets denote summation with respect to h.

Equations (5) can be written down easily, since the coefficients w_{jki} appear in the same order as in (2). The equations corresponding to each i form a complete set which can be solved independently of those for other values of i. The solution can be expressed

$$\begin{aligned} V_{ji} &= A_{1ji}[\lambda_h F^h_{1i}] + A_{2ji}[\lambda_h F^h_{2i}] + \cdots + A_{qji}[\lambda_h F^h_{qi}] \\ &= \left[\lambda_h \sum_{k=1}^{q} A_{kji} F^h_{ki}\right] \end{aligned} \tag{7}$$

where A_{kji} is $(-1)^{k+j}$ times the minor corresponding to w_{kji}, divided by the principal determinant. By symmetry $A_{kji} = A_{jki}$.

The V's in (4) are to be replaced by their values from (7) and the coefficients of the λ's collected. To facilitate this let

$$L_{jk} = \sum_{i=1}^{n} L_{jki}$$

where

$$L_{jki} = \sum_{s=1}^{q} \sum_{r=1}^{q} A_{rsi} F^j_{ri} F^k_{si}.$$

Each L_{jki} can be written down easily from the corresponding Q_i as written in (2): in each term $w_{rsi} V_{ri} V_{si}$ replace w_{rsi} by A_{rsi}, V_{ri} by F^j_{ri}, and V_{si} by F^k_{si}. It is important to preserve the order of the subscripts of the V's in (2), and to treat the diagonal terms $w_{rri} V^2_{ri}$ as though written $w_{rri} V_{ri} V_{ri}$. It is seen that $L_{jki} = L_{kji}$, and $L_{jk} = L_{kj}$. Then the substitution from (7) into (4) gives

$$\sum_{j=1}^{p} L_{jk} \lambda_j + \sum_{l=1}^{r} F^k_l v_l = F^k_0 \qquad k = 1, 2, \ldots, p. \tag{8}$$

Equations (8), with (6), are formally identical with those of the least squares procedure which are called by Deming the "general normal equations", and they can be written schematically in the same manner. The further procedure is identical with that for least squares, involving solution of the general normal equations for the λ's and v's, substitution of the values of the λ's into (7) to obtain the V's, and then adjustment of the observations by use of the V's, and adjustment of the provisional values of the parameters by use of the v's.

A word of appreciation is due Dr. O. W. Richards of The Spencer Lens Company for calling this problem to my attention, and for encouragement in the carrying out of the solution.

REFERENCES

[1] Joseph Berkson, "Growth changes in physical correlation--height, weight, and chest-circumference, males," *Human Biology*, Vol. 1(1929), pp. 462-502.

[2] W Edwards Deming, "The application of least squares," *Phil. Mag.* 7th Ser., Vol. 11(1931), pp 146-158

[3] W. Edwards Deming, "On the application of least squares," *Phil. Mag.* 7th Ser , Vol. 17(1934), pp. 804-829.

[4] W. Edwards Deming, *Proc. Phys. Soc. Lond.*, Vol. 47(1935), pp. 92-106.

[5] W. Edwards Deming, *Some Notes on Least Squares*, Graduate School, Dept. of Agric., Washington.

THE ESTIMATION OF A QUOTIENT WHEN THE DENOMINATOR IS NORMALLY DISTRIBUTED

By Robert D. Gordon

Scripps Institution of Oceanography, La Jolla, Calif.

1. Introduction. In an oceanographic investigation we have to deal with a time series consisting of single pairs of observed values x, y, of two independent stochastic variables, whose true (mean) values we shall denote respectively by a, b. Of interest is the corresponding time series of quotients (b/a), which it is required to estimate from the observations x, y. Both x and y are approximately normally distributed about their mean values a, b with rather large variances σ_x^2, σ_y^2 which can be estimated. It is easily possible for x to vanish or even to be of opposite sign to a, although a cannot itself vanish. The required estimates of (b/a) should have the property that they can be numerically integrated, i.e. that an arbitrary sum of such estimates shall equal the corresponding estimate of the true sum.

Let us define a function $\gamma(x)$ to have the property that its mathematical expectation $E\{\gamma(x)\}$ is exactly $1/a$, where $a = E(x)$. If such a function exists we shall have

$$E\{y\cdot\gamma(x)\} = E(y)\cdot E\{\gamma(x)\} = b\cdot(1/a) = b/a \tag{1}$$

so that $y\cdot\gamma(x)$ will be an estimate of b/a which has the required property: namely such estimates can be added, and we have

$$E\{y_1\gamma(x_1) + y_2\gamma(x_2)\} = E\{y_1\gamma(x_1)\} + E\{y_2\gamma(x_2)\} = b_1/a_1 + b_2/a_2$$

as required. It turns out that if x is normally distributed with non-zero mean such a function $\gamma(x)$ does exist, and is given by the formula

$$\gamma(x) = \frac{1}{\sigma_x}\exp(x^2/2\sigma_x^2)\int_{x/\sigma_x}^{\infty} e^{-t^2/2}\,dt = \frac{1}{\sigma_x}R_{x/\sigma_x} \tag{2}$$

where R_u is the "ratio of the area to the bounding ordinate" which is tabulated by J. P. Mills,[1] also in Pearson's tables.[2] Equation (2) holds if a is positive; if a is negative the integration should extend over $(x/\sigma_x, -\infty)$. It is easy to verify that

$$E(\gamma(x)) = \frac{1}{\sqrt{2\pi}\,\sigma_x}\int_{-\infty}^{\infty}\gamma(x)\exp\left(-\frac{(x-a)^2}{2\sigma_x^2}\right)dx = \frac{1}{a} \tag{3}$$

by direct substitution from (2).

[1] J. P. Mills, "Table of ratio: area to bounding ordinate, for any portion of the normal curve," *Biometrika*, Vol. 18 (1926), pp. 395–400.

[2] Karl Pearson, *Tables for Statisticians and Biometricians*, part II, table III, Cambridge Univ. Press.

2. The law of large numbers for $\gamma(x)$. The function $\gamma(x)$ defined by (2) has mean value $1/a$ as required, but its second moment (hence variance) does not exist, as may readily be verified. By a theorem of Khinchine[3] however, its values satisfy a law of large numbers. It will be of interest to inquire about the "strength" of this law of large numbers for $\gamma(x)$. Namely, given a positive number ϵ, how many "observations" (independent estimates) $\gamma(x)$ will suffice to guarantee probabilities of .50, .90, .95, etc. for the following inequality to hold

$$\left| \frac{\gamma(x_1) + \gamma(x_2) + \cdots + \gamma(x_n)}{n} - \frac{1}{a} \right| < \epsilon \tag{4}$$

where n is the number of "observations."

In order to arrive at a rough answer to this question we have made use of certain inequalities due to Tshebysheff (Tshebysheff's "method of moments", cf. Uspensky[4]). Let u be an arbitrary stochastic variable whose distribution has moments of the first and second order which are known. Denote by m its first moment, by σ^2 its variance, then it results from Tshebysheff's theory that the probability $P(u_1, u_2)$ for a value of u to lie between u_1 and u_2 (i.e. $u_1 \leqq u \leqq u_2$) satisfies the inequality

$$P(u_1, u_2) > 1 - \frac{\sigma^2}{(u_1 - m)^2 + \sigma^2} - \frac{\sigma^2}{(u_2 - m)^2 + \sigma^2}. \tag{5}$$

This inequality is independent of the values, or even the existence, of further moments of the u-distribution beyond the second, and depends only on the condition that the cumulant of the distribution function shall have at least three "points of increase."

Although $\gamma(x)$ does not have a second moment, a second moment does exist for those values of $\gamma(x)$ which correspond to $x \geqq -\theta > -\infty$, where θ is an arbitrary number, positive or negative. If we can estimate the first two moments of $\gamma(x) \sim 1/x$ corresponding to a given value of θ, then for a given number n of observations we need only to divide the corresponding variance by n to obtain σ^2 in (5), then multiply (5) by the nth power of the (normal) probability for the inequality $x \geqq -\theta$, in order to obtain a lower bound for the probability of the inequality (4). θ is to be determined so as to yield a maximum result.

The first moment m_1 of $\gamma(x)$ for values of $x \geqq -\theta$ is easily computed, and is given by the formula

$$\sigma_x m_1(\theta) = \frac{\sigma_x}{a}\left\{1 - \frac{R_{-\theta/\sigma_x}}{R_{-(\theta+a)/\sigma_x}}\right\}. \tag{6}$$

---

[3] J. V. Uspensky, *Introduction to Mathematical Probability*, pp. 195, McGraw-Hill (1937).
[4] J. V. Uspensky, l.c. pp. 365 ff.

The second moment is harder to compute, but if we place

$$\phi(\theta) \quad K\cdot(m_2 - m_1^2) = \frac{1}{\sqrt{2\pi}\,\sigma_x}\int_\theta^\infty [\gamma(x)]^2 \exp\left(-\frac{(x-a)^2}{2\sigma_x^2}\right)dx$$

$$(7) \qquad - \frac{\left[\int_\theta^\infty \gamma(x)\exp\left(-\frac{(x-a)^2}{2\sigma_x^2}\right)dx\right]^2}{\sqrt{2\pi}\,\sigma_x\int_\theta^\infty \exp\left(-\frac{(x-a)^2}{2\sigma_x^2}\right)dx}$$

where

$$K = \frac{1}{\sqrt{2\pi}\,\sigma_x}\int_{-\theta}^\infty \exp\left(-\frac{(x-a)^2}{2\sigma_x^2}\right)dx = \frac{1}{\sqrt{2\pi}}\int_{-(\theta+a)/\sigma_x}^\infty e^{-t^2/2}\,dt$$

we easily obtain the relationship

$$(8) \quad \phi'(\theta) = \frac{1}{\sqrt{2\pi}\,\sigma_x^3}\exp\left(-\frac{(\theta+a)^2}{2\sigma_x^2}\right)\left\{R_{(-\theta/\sigma_x)} - \frac{\sigma_x}{a}\left(1 - \frac{R_{-\theta/\sigma_x}}{R_{-(\theta+a)/\sigma_x}}\right)\right\}^2.$$

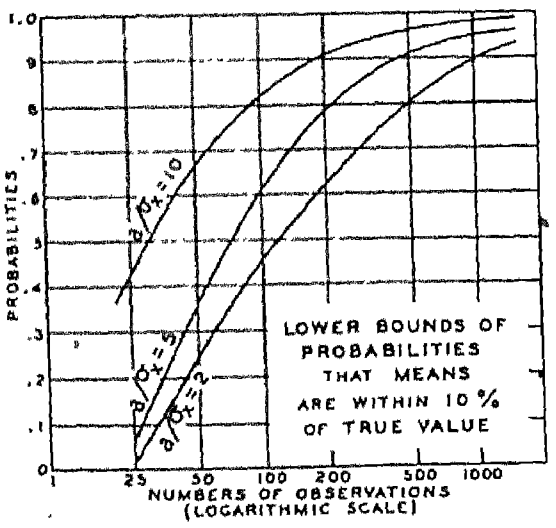


FIG. 1

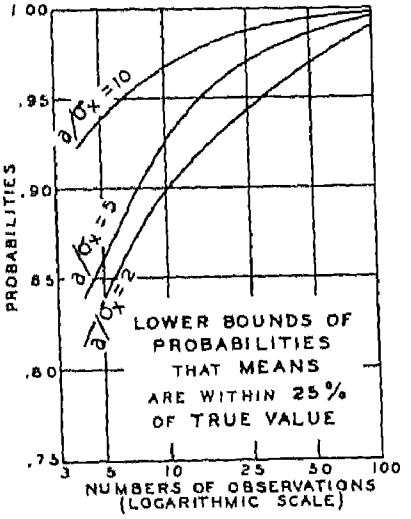


FIG. 2

From (7), using a table of the probability integral, it can be verified that $\phi(-a - 3\sigma_x) \ll 0.001$. Assume, therefore, as a boundary condition $\phi(-a - 3\sigma_x) = 0$ then (8) can be integrated graphically or numerically. It is by this means that the curves shown in Figs. 1 and 2 were determined. Computations were also attempted for $a/\sigma_x = \frac{1}{2}$, $a/\sigma_x = 1$, but it was not possible to obtain significant results: it would be necessary in these cases to take more than two moments into account, which would lead to hopeless complications. In these figures the ordinates represent probabilities for an observation to fall between *.90a* and *1.11a* (Figure 1), and between *.75a* and *1.33a* (Figure 2), respectively.

3. Two practical formulas for computations. It seems worthwhile to note here two simple formulas in connection with Mills' ratio (2) which will be useful for computations. The first is the obvious relationship

$$R_{-u} = \sqrt{2\pi}\, e^{u^2/2} - R_u = 1/z - R_u \tag{9}$$

in the notation of Pearson's tables. The second applies to large values of x, and may be written

$$\frac{x}{x^2 + \sigma_x^2} < \gamma(x) = \frac{1}{\sigma_x} R_{x/\sigma_x} < \frac{1}{x} \tag{10}$$

(10) is true for $x > 0$, and can be proved by means of the differential equation which $\gamma(x)$ satisfies.

4. Remarks. The estimate $\gamma(x)$ has the following inadequacy: If only a single observation x is known, then it is unknown whether a is of like or unlike sign compared to x. It turns out then that the mathematical expectation for the value of $\gamma(x)$ vanishes identically. This difficulty of course disappears if more than one observation is available. Methods of avoiding this difficulty for time series, e.g. by noting relative frequencies for observations separated by 1, 2, 3 etc. intervals to agree in sign, will be discussed elsewhere in connection with practical applications.

It may be worthwhile to note that Geary[6] developed certain characteristics of the distribution of a quotient, which however are not adapted to our purposes.

NOTE ON CONFIDENCE LIMITS FOR CONTINUOUS DISTRIBUTION FUNCTIONS

By A. Wald* and J. Wolfowitz

In a recent paper [1] we discussed the following problem: Let X be a stochastic variable with the cumulative distribution function $f(x)$, about which nothing is known except that it is continuous. Let $x_1, \cdots, x_n$ be n independent, random observations on X. The question is to give confidence limits for $f(x)$. We gave a theoretical solution when the confidence set is a particularly simple and important one, a "belt."

A particularly simple and expedient way from the practical point of view is to construct these belts of uniform thickness ([1], p. 115, equation 50). If the appropriate tables, as mentioned in our paper, were available, the construction of confidence limits, no matter how large the size of the sample, would be immediate.

Our formulas (11), (16), (19), (27) and (29) are not very practical for computation, particularly when the samples are large. We have recently learned that

[6] Geary, R. C., "The Frequency Distribution of a Quotient," *Jour. Roy. Stat. Soc.*, Vol 93 (1930), pp. 442–446.

* Columbia University, New York City.

there exists a result by Kolmogoroff [2], generalized by Smirnoff [3],[1] which for large samples gives an easy method for constructing tables, i.e. of finding α when c and n are given (all notations as in [1]). The result of Kolmogoroff-Smirnoff is:

Let $c = \lambda/\sqrt{n}$. Then for any fixed $\lambda > 0$,

$$\lim_{n=\infty} \bar{P} = \lim_{n=\infty} \underline{P} = 1 - e^{-2\lambda^2}$$

$$\lim_{n=\infty} P = 1 - 2\sum_{m=1}^{\infty} (-1)^{(m-1)} e^{-2m^2\lambda^2}.$$

This series converges very rapidly.

REFERENCES

[1] Wald and Wolfowitz, "Confidence limits for continuous distribution functions," *Annals of Math. Stat.*, Vol. 10(1939), pp. 105-118.

[2] A. Kolmogoroff, "Sulla determinazione empirica di una leggi di distribuzione," *Giornale dell'Instituto Italiana degli Attuari*, Vol. 11(1933)

[3] N. Smirnoff, "Sur les ecarts de la courbe de distribution empirique," *Recueil Mathematique (Mathematicheski Sbornik)*, New series, Vol. 6(48)(1939), pp. 3-26.

[1] In the French résumé of Smirnoff's article, on page 26, due to a typographical error this formula is given with a factor $(-1)^m$ instead of the correct factor $(-1)^{m-1}$. The correct result follows from equation (112), page 23, of the Russian text when t is set equal to zero.

REPORT OF THE CHICAGO MEETING OF THE INSTITUTE

The Sixth Annual Meeting of the Institute of Mathematical Statistics was held at the Stevens Hotel, Chicago, Thursday to Saturday, December 26 to 28, 1940 in conjunction with the meetings of the American Statistical Association, the Econometric Society, and the American Marketing Association. The following fifty members of the Institute attended the meeting:

H. E Arnold, C S Barrett, A G. Brooks, R W Burgess, A. G Clark, A. C Cohen, Jr., W G Cochran, A. T Craig, C C. Craig, B. B. Day, W E. Deming, J. L. Doob, P. S Dwyer, Churchill Eisenhart, J W Fertig, P. G Fox, Hilda Geiringer, E J. Gumbel, Myron Headingsfield, Harold Hotelling, Leo Katz, J. F. Kenney, L F Knudsen, Alma Kohl, T. Koopmans, D. H Leavens, Ida Levin, G A Lundberg, S. N. Lyttle, W G Madow, Ralph Mansfield, G. F. T Mayer, J. R. Miner, E C. Molina, C. R Mummery, J I Northam, E G. Olds, P. S. Olmstead, A L. O'Toole, J A. Pierce, Wilhelm Reitz, P R Rider, M M. Sandomire, L. W. Shaw, W A. Shewhart, F. F Stephan, S. A. Stouffer, A. G Swanson, S S. Wilks, M O. Woodbury

The opening session, on Thursday afternoon, was devoted to contributed papers in probability and statistical methodology. The Chairman was Professor S. S. Wilks of Princeton University, and the following papers were presented:

1. *On the Calculation of the Probability Integral on Non-Central t and an Application.*
   C. C Craig, University of Michigan.
2. *Effective Methods of Graduation*
   Max Sasuly, Office of the Actuary, Social Security Board.
3. *On Some New Results in the Sampling of Discrete Random Variables.*
   William G Madow, Bureau of the Census.
4. *On the Use of Inverse Probability in Sample Inspection*
   W. Edwards Deming and W G Madow, Bureau of the Census
5. *On a Convergent Iterative Procedure for Adjusting a Sample Frequency Table when Some of the Marginal Totals are Known*
   F F. Stephan, Cornell University, and W. Edwards Deming, Bureau of the Census.
6. *The Return Period of Flood Flows.*
   E J. Gumbel, New School for Social Research, New York City.
7. *A Note on the Power of a Sign Test.*
   W M. Stewart, University of Wisconsin.
8. *A New Explanation of Non-Normal Dispersion.*
   Hilda Geiringer, Bryn Mawr College.

Abstracts of these papers follow this report.

On Friday morning a session was held jointly with the American Marketing Association on *The Theory and Application of Representative Sampling.* Under the chairmanship of Professor Theodore H. Brown of Harvard University, the following papers were presented:

1 *Background and Method.*
   F. F. Stephan, Cornell University.

2. *Application to Marketing Problems.*
   Archibald M. Crossley, New York City.
3. *Application to Agricultural Problems.*
   Arnold J. King, Iowa State College

The afternoon session on Friday was held jointly with the American Statistical Association and Econometric Society on *The Analysis of Variance.* The chair was held by Professor P. R. Rider of Washington University and the following papers were presented:

1. *The Relation Between the Design of an Experiment and the Analysis of Variance.*
   A. E. Brandt, Soil Conservation Service.
2. *The Underlying Principles of the Analysis of Variance and Associated Tests of Significance.*
   Churchill Eisenhart, University of Wisconsin.
3. *The Applications of the Analysis of Variance to Non-Orthogonal Data.*
   W. G. Cochran, Iowa State College

Discussion:
   Gertrude M. Cox, North Carolina State College.
   John F. Kenney, University of Wisconsin Extension Division.
   W. Edwards Deming, Bureau of the Census

On Saturday morning and afternoon, sessions were held with the American Statistical Association on *Collection and Use of Statistics for Quality Control in National Defense Industries.* At the morning session the following papers were given, with Dr. C. W. Gates of the Western Electric Company in the chair:

1. *Report on the Quality Control Program of the American Standards Association.*
   John Gaillard, Western Electric Company.
2. *Sample Verification in the Administration of the Population Census.*
   W. Edwards Deming, Bureau of the Census.
3. *The Importance of the Statistical Viewpoint in High Production Manufacturing.*
   P. L. Alger, General Electric Company. (Read by C. Eisenhart)
4. *On the Initiation of Statistical Methods for Quality Control in Industry.*
   Leslie E. Simon, Aberdeen Proving Ground.

At the afternoon session the following papers were presented under the chairmanship of Dr. John Johnston of the United States Steel Corporation:

1. *The Place of Statistical Analysis in Ferrous Metallurgy.*
   E. M. Schrock, Jones and Laughlin Steel Corporation.
2. *Statistical Methods in the Production and Inspection of Cast Iron Pipe.*
   J. T. MacKenzie, American Cast Iron Pipe Company.
3. *Applications of Statistical Methods to Metallurgy.*
   R. B. Mears, Aluminum Company of America.

Discussion:
   Churchill Eisenhart, University of Wisconsin.

The annual business meeting of the Institute was held on Thursday afternoon after the session on probability and statistical methodology, with the President presiding.

The Secretary-Treasurer read the financial report for 1940.

The Editor of the *Annals of Mathematical Statistics* reported on the progress of the *Annals* during 1940. It was stated that manuscripts worthy of publication were now being submitted at a rate that would justify the publication of a 500-page annual volume. To make this amount of publication self-supporting upon the expiration of the Rockefeller grant in June, 1941, it was pointed out that another 150 new subscriptions would have to be obtained during 1941. Judging from the rate at which subscriptions had been coming in during the past two years such an increase was considered entirely feasible with the cooperation of the members of the Institute. Various methods of effecting this increase were discussed at the meeting and suggested for the consideration of the Board of Directors.

On behalf of the Board of Directors the President made the following report:

1. The Report of the War Preparedness Committee, approved in preliminary form at the Hanover meeting, had been preprinted and some of the preprints had already been distributed.

2. Arrangements had been made with the Executive Officer of the National Roster of Scientific and Specialized Personnel to send the statistics check list to all members of the Institute who are not members of the American Statistical Association.

3. That preprints of the pamphlet on *The Teaching of Statistics*, including an address by Professor Harold Hotelling, discussion by Dr. W. E. Deming and the resolutions on the teaching of statistics adopted by the Institute at the Dartmouth meeting had been produced and distributed.

4. That application[1] had been made to the Executive Committee of the American Association for the Advancement of Science through the Permanent Secretary for admission to the status of an affiliated society in the Association.

It was announced that through the annual election, carried out by mail ballot, the following officers were elected for 1941 (all names being those proposed by the Nominating Committee):

President: Professor Harold Hotelling
Vice-Presidents: Professor A. T. Craig
Professor H. C. Carver
Secretary-Treasurer: Professor E. G. Olds

The annual luncheon was held at noon on Friday with the President-Elect presiding. Short talks were made by Dr. E. J. Gumbel, Dr. T. Koopmans and Professor S. S. Wilks, while the annual luncheon address was delivered by Professor P. R. Rider.

P. R. RIDER,
Secretary-Treasurer.

[1] This application was approved by the Executive Committee of the A.A.A.S. at its December 1940 meeting.

ABSTRACTS OF PAPERS

(Presented on December 26, 1940, at the Chicago meeting of the Institute)

On the Calculation of the Probability Integral on Non-Central t and an Application. C. C. CRAIG, University of Michigan.

It seems not to have been noted that the probability integral for non-central t can be calculated by means of an infinite series in incomplete β-functions which converges rapidly for small samples. The application here considered is to a test based on the randomization principle which is the subject of E. J. G. Pitman's paper, *Significance tests which may be applied to samples from any populations* (*Roy. Stat. Soc. Jour.*, Vol. 4 (1937), pp. 119–130). In case the samples come from normal populations with equal variance but with unequal means, the chance that the hypothesis of equal population means will be accepted on this test is given by this probability integral which is evaluated in some illustrative numerical examples.

On Some New Results in the Sampling of Discrete Random Variables. WILLIAM G. MADOW, Bureau of the Census.

Many statistical tables may be regarded as the result of subsampling finite populations classified into $r \times s \times \cdots$ tables. The main aim of this paper is to derive the associated statistical theory including both the finite and limiting distributions. After evaluating the fundamental distributions and the moments it is shown that under certain conditions, the limiting distribution is multinomial, while under other conditions the limiting distribution is multivariate normal. These results are then applied to determine the adequate size of sample, and the sampling proportions from various strata.

On the Use of Inverse Probability in Sample Inspection. W. EDWARDS DEMING and WILLIAM G. MADOW, Bureau of the Census.

The theory of inspection by sampling is abstractly equivalent to one part of the theory of subsampling. The theory of subsampling finite populations is considered in this paper in order to investigate the differences that occur when the methods of fiducial inference and inverse probability are used, particularly in regard to determining the adequate size of sample. In sample inspection, the prior distribution of failures is almost always known, at least approximately. In using any system of sample inspection, a number of failures will pass undetected. On the basis of certain prior distributions of failures, distributions are derived for the number and percent of failures remaining after each of several different possible systems of sample inspection has been applied. Formulas giving the cost of partial inspection are used together with these distributions in order to determine methods of sample inspection having various desired properties.

On a Convergent Iterative Procedure for Adjusting a Sample Frequency Table When Some of the Marginal Totals are Known. FREDERICK F. STEPHAN, Cornell University and W. EDWARDS DEMING, Bureau of the Census.

The 5 per cent sample taken with the 1940 Population Census presents an interesting problem of estimation in which the estimates are connected by equations of condition.

These equations arise from the fact that certain sums of estimates derived from the sample should equal the corresponding frequencies derived from the tabulations of the census enumeration, i e the distribution of each of several variables may be known but their joint distribution may only be estimated from a cross tabulation of the data furnished by the sample. The adjustment of the sample estimates is accomplished by the principle of least squares and an outline of the various types of conditions for two and three variables is presented. The solution of the normal and condition equations is tedious when hundreds of sets of estimates must be adjusted but a simple iterative procedure is available (see *Annals of Math. Stat.*, Vol 11 (1940), pp. 427 444)

The Return Period of Flood Flows. E. J. GUMBEL, New School for Social Research (N. Y.)

For any statistical variable the return period is defined as the mean number of trials necessary in order that a certain value of the variable or a greater one returns. The return period is a theoretical statistical function such as the distribution or the probability. In hydraulics the corresponding observed values are the recurrence and exceedance intervals.

The main thesis is that *the flood flows are the largest values of flows* which have to be considered as *unlimited* variables. The method of return periods applied to the largest values leads without further assumptions to a formula which gives the return period $f(x)$ of a flood superior to x, and at the same time the most probable flood to be reached not at a certain time, but within a certain period. This formula contains only two constants, which are linear functions of the mean annual flood and the standard deviation. Fuller's formula turns out to be an asymptotic expression of my formula.

This method applied to the Connecticut, Columbia, Merrimack, Cumberland, Tennessee and Mississippi rivers shows a very good fit between theory and observation, superior to the methods applied heretofore.

A Note on the Power of the Sign Test. W. M. STEWART, University of Wisconsin.

Let us consider a set of N non-zero differences, of which x are positive and $N - x$ are negative, and suppose that the hypothesis tested, H_0, implies in independent sampling that x will be distributed about an expected value of $N/2$ in accordance with the binomial $(\frac{1}{2} + \frac{1}{2})^N$. As a quick test of H_0, we may choose to test the hypothesis h_0 that x has the above probability distribution. Defining r to be the smaller of x and $N - x$, the test consists in rejecting h_0 and therefore H_0 whenever $r \leq r(\epsilon, N)$, where $r(\epsilon, N)$ is determined by N and the significance level ϵ.

In applying such a test it is of interest to know how frequently it will lead to a rejection of H_0 when H_0 is false and the actual situation H implies that the probability law of x is $(q + p)^N$, with $p \neq \frac{1}{2}$, thereby indicating an expectation of an unequal number of + and − differences. The probability of rejecting H_0 when H_1 implying $p = p_1$ is true, is termed the *power* of the test of H_0 relative to the alternative H_1.

A table is given for the 5% significance level ($\epsilon = .05$) showing the minimum value of N for which the power of the test relative to $p = p_1$ exceeds β for values of β from .05 to .95 at intervals of 05; and for p_1 from .60 to .95 (and thereby for p_1 from .40 to .05) at intervals of .05. The case of $\beta > .99$ is also considered for these values of p_1

A New Explanation of Non-Normal Dispersion. HILDA GEIRINGER, Bryn Mawr College.

The starting point of the *Lexis theory* consists in this fact: It is to be expected, on the average, that two expressions Σ and Σ' which can be computed from the results of $m \cdot n$ observations are equal, provided that the corresponding $m \cdot n$ chance variables $x_{\mu\nu}$ are

equally and independently distributed. Let a_μ be the average $a_\mu = 1/n \sum_{\nu=1}^{n} x_{\mu\nu}$ and a the average of the a_μ ($\mu = 1, \cdots, m$). Then

$$\sum' = \frac{m}{mn-1} s'^2 = \frac{m}{mn-1} \frac{\sum_\mu \sum_\nu (x_{\mu\nu} - a)^2}{mn}$$

$$\sum = \frac{m}{m-1} s^2 = \frac{m}{m-1} \frac{\sum_\mu (a_\mu - a)^2}{m}.$$

We see, however, that rows and columns do not play the same role here because Σ depends only on the a_μ, the average values of the rows. If the observed value of Σ happens to be larger (smaller) than the value of Σ', we speak of supernormal (subnormal) dispersion. It is well known that supernormal dispersion can be explained by assuming that the m n theoretical populations are only equal "by rows" but not by columns (there are m different distributions); in the same way one can explain the case of subnormal dispersion by admitting that the distributions are equal "by columns," but not by rows.

Another explanation which may sometimes seem more plausible is the following: All the $m \cdot n$ distributions are supposed to be equal, but we *omit the assumption of mutual independence*. Then one can prove that the supernormal or subnormal dispersion corresponds respectively to an appropriately defined "positive" or "negative correlation." The fact that normal dispersion occurs rather rarely in social questions is then reflected by the idea that social phenomena are in fact not independent of each other but are usually only assumed so for the purpose of simplicity. In that way the more frequent occurrence of supernormal dispersion likewise finds an adequate explanation.

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

PAGE

The Cyclic Effects of Linear Graduations Persisting in the Differences of the Graduated Values. EDWARD L. DODD 127

On the Distribution of Wilks' Statistic for Testing the Independence of Several Groups of Variates. A. WALD and R. J. BROOKNER. 137

The Mean Square Successive Difference. J. VON NEUMANN, R. H. KENT, H. R. BELLINSON, and B. I. HART 153

The Return Period of Flood Flows. E. J. GUMBEL 163

On the Foundations of Probability and Statistics. R. VON MISES. 191

Probability as Measure. J. L. DOOB 206

Discussion of Papers on Probability Theory. R. VON MISES and J. L. DOOB 215

Continued Fractions for the Incomplete Beta Function. LEO A. AROIAN 218

Notes:

Note on the Distribution of Non-central t with an Application. CECIL C. CRAIG 224

Note on an Application of Runs to Quality Control Charts. FREDERICK MOSTELLER 228

Test of Homogeneity for Normal Populations. G. A. BAKER 233

A Note on the Power of the Sign Test. W. MAC STEWART 236

Moments of the Ratio of the Mean Square Successive Difference to the Mean Square Difference in Samples from a Normal Universe. J. D. WILLIAMS 239

Vol. XII, No. 2 — June, 1941

THE
'HEMA — 'ATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG J. NEYMAN

WITH THE COÖPERATION OF

H. C. CARVER
H. CRAMÉR
W. E. DEMING
G. DARMOIS
R. A. FISHER
T. C. FRY
H. HOTELLING
R. VON MISES
E. S. PEARSON
H. L. RIETZ
W. A. SHEWHART

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I-IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, Inc.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

that is to make the graph a perfect zig-zag where positive and negative values of z_i alternate. A set of differencings following a set of summings may bring the cycle length from some fairly large number back to about 3, and thus restore something like the original chaotic appearance in the graph.

In dealing with the foregoing $\Phi(x)$ or $\phi(x)$ in (2), it was not assumed that the distribution be normal. But, in what follows, it will be assumed that

$$\phi(x) = \frac{1}{\sigma(2\pi)^{1/2}} e^{-(x-\mu)^2/2\sigma^2}; \tag{3}$$

and, for convenience, μ will be taken as zero—that is, the data will be supposed given as deviations from their theoretic mean. Actually, the data used by Slutzky and the data I have used belong to a rectangular distribution, as noted in my former paper. Nevertheless the close agreement between actual and expected results seems to indicate [3, p. 263] that the theory is in general applicable. It is well known that averaging of observations from non-normal distributions may lead rather quickly to an approximately normal distribution.

Given n real numbers, $a_1, a_2, \cdots, a_n$, let

$$y_i = a_1x_i + a_2x_{i+1} + \cdots + a_nx_{i+n-1}; \qquad i = 1, 2, 3, \cdots. \tag{4}$$

Then y_i is the moving *sum* if each $a_r = 1$. Slutzky takes $j = i$ or $j = i + n - 1$ Again, y_i is the moving *average* if each $a_r = 1/n$. For graduation in general, the condition $\Sigma a_r = 1$ is imposed; and usually $j = i + (n + 1)/2$ If n is odd, y_j is thus associated with the middle x.

Under the assumption that the x's are independent and normally distributed about mean zero, with constant variance, I have proven [3, p. 256]: The probability that for any specified j, $y_{j-1} < 0$, and $y_j > 0$ is given by $P = \theta/360°$, where

$$\cos\theta = \sum_{r=1}^{n-1} a_r a_{r+1} \Big/ \sum_{r=1}^{r=n} a_r^2. \tag{5}$$

The expected relative frequency of up-crossings of the graph of the y's through the zero base line is then $\theta/360°$. That is: $\theta/360°$ is the expected relative frequency of a change in the sign of y from $-$ to $+$; also, of a change in sign from $+$ to $-$.

But, as $\Delta y_i = y_{i+1} - y_i$, it follows that

$$\Delta y_i = b_1x_i + b_2x_{i+1} + \cdots + b_nx_{i+n-1} + b_{n+1}x_{i+n}, \tag{6}$$

where

$$b_1 = -a_1, \quad b_{n+1} = a_n, \quad b_r = a_{r-1} - a_r, \qquad r = 2, 3, \cdots, n-1 \tag{7}$$

and since a maximum for the y's at y_i occurs when $\Delta y_{i-1} > 0$, $\Delta y_i < 0$, it follows that the theoretic frequency therefor is $\theta'/360°$, where

$$\cos\theta' = \sum_{r=1}^{n} b_r b_{r+1} \Big/ \sum_{r=1}^{n+1} b_r^2. \tag{8}$$

THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG J. NEYMAN

WITH THE COÖPERATION OF

H. C. CARVER
H. CRAMÉR
W. E. DEMING
C. DARMOIS
R. A. FISHER
T. C. FRY
H. HOTELLING
R. VON MISES
E. S. PEARSON
H. L. RIETZ
W. A. SHEWHART

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I-IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

that is to make the graph a perfect zig-zag where positive and negative values of z_i alternate. A set of differencings following a set of summings may bring the cycle length from some fairly large number back to about 3, and thus restore something like the original chaotic appearance in the graph.

In dealing with the foregoing $\Phi(x)$ or $\phi(x)$ in (2), it was not assumed that the distribution be normal. But, in what follows, it will be assumed that

$$\phi(x) = \frac{1}{\sigma(2\pi)^{1/2}} e^{-(x-\mu)^2/2\sigma^2}; \tag{3}$$

and, for convenience, μ will be taken as zero—that is, the data will be supposed given as deviations from their theoretic mean. Actually, the data used by Slutzky and the data I have used belong to a rectangular distribution, as noted in my former paper. Nevertheless the close agreement between actual and expected results seems to indicate [3, p. 263] that the theory is in general applicable. It is well known that averaging of observations from non-normal distributions may lead rather quickly to an approximately normal distribution.

Given n real numbers, $a_1, a_2, \cdots, a_n$, let

$$y_i = a_1x_i + a_2x_{i+1} + \cdots + a_nx_{i+n-1}, \qquad i = 1, 2, 3, \cdots \tag{4}$$

Then y_i is the moving *sum* if each $a_r = 1$. Slutzky takes $j = i$ or $j = i + n - 1$ Again, y_i is the moving *average* if each $a_r = 1/n$. For graduation in general, the condition $\Sigma a_r = 1$ is imposed; and usually $j = i + (n + 1)/2$. If n is odd, y_i is thus associated with the middle x.

Under the assumption that the x's are independent and normally distributed about mean zero, with constant variance, I have proven [3, p. 256]: The probability that for any specified j, $y_{j-1} < 0$, and $y_j > 0$ is given by $P = \theta/360°$, where

$$\cos\theta = \sum_{r=1}^{r=n-1} a_r a_{r+1} \Big/ \sum_{r=1}^{r=n} a_r^2. \tag{5}$$

The expected relative frequency of up-crossings of the graph of the y's through the zero base line is then $\theta/360°$. That is: $\theta/360°$ is the expected relative frequency of a change in the sign of y from $-$ to $+$; also, of a change in sign from $+$ to $-$.

But, as $\Delta y_i = y_{i+1} - y_i$, it follows that

$$\Delta y_i = b_1x_i + b_2x_{i+1} + \cdots + b_nx_{i+n-1} + b_{n+1}x_{i+n}, \tag{6}$$

where

$$b_1 = -a_1, \quad b_{n+1} = a_n, \quad b_r = a_{r-1} - a_r, \qquad r = 2, 3, \cdots, n-1 \tag{7}$$

and since a maximum for the y's at y_i occurs when $\Delta y_{i-1} > 0$, $\Delta y_i < 0$, it follows that the theoretic frequency therefor is $\theta'/360°$, where

$$\cos\theta' = \sum_{r=1}^{n} b_r b_{r+1} \Big/ \sum_{r=1}^{n+1} b_r^2. \tag{8}$$

In a similar manner, by using *second* differences, we get the expected relative frequency $\theta''/360°$ for inflexional points, in specified direction. Moreover, $\theta \leqq \theta' \leqq \theta'' \leqq \cdots \leqq 180°$; since inflections must be at least as frequent as maxima, etc.

If the foregoing formulas are applied to the identical "graduation" $y_i = x_i$, then $\cos\theta = 0$, $\cos\theta' = -1/2$, $\cos\theta'' = -2/3$. In fact,

$$\cos\theta^{(t)} = -t/(t+1). \tag{9}$$

This follows from the fact that the b's and similar coefficients are the binomial coefficients; and

$$\sum_{r=0}^{t} {}_tC_r^2 = {}_{2t}C_t; \qquad \sum_{r=0}^{t-1} {}_tC_r \cdot {}_tC_{r+1} = {}_{2t}C_{t-1}. \tag{10}$$

Thus repeated differencing leads toward the perfect zig-zag. An extension of this feature will be taken up in the next section.

3. Repeated summing and differencing. To indicate the *result* of the summing of n consecutive numbers in a sequence, I shall use the notation 1^n. And the difference $\Delta y_i = y_{i+1} - y_i$ will be indicated by $-1, 0^{n-1}, 1$. Thus if $n = 3$, 1^3 and $-1, 0^2, 1$ will stand respectively for

$$y_i = x_{i-1} + x_i + x_{i+1}; \qquad \Delta y_i = -x_{i-1} + 0x_i + 0x_{i+1} + x_{i+2}. \tag{11}$$

If, now, $z_i = y_{i-1} + y_i + y_{i+1}$, then

$$z_i = x_{i-2} + 2x_{i-1} + 3x_i + 2x_{i+1} + x_{i+2}. \tag{12}$$

Since (n) is often used to indicate the *operation* of summing n consecutive numbers, we may write

$$(3)^2 = 1, 2, 3, 2, 1; \qquad (n)^2 = 1, 2, \cdots, (n-1), n, (n-1), \cdots, 2, 1. \tag{13}$$

Then, for $n \geq 2$,

$$\Delta(n)^2 = -1^n, 1^n; \qquad \Delta^2(n)^2 = 1, 0^{n-1}, -2, 0^{n-1}, 1. \tag{14}$$

And, since the operations of summing and differencing are commutative, we are lead to

$$F_n^k = (-1)^k\Delta^k(n)^k = {}_kC_0, 0^{n-1}, -{}_kC_1, 0^{n-1}, {}_kC_2, 0^{n-1}, \cdots, (-1)^k{}_kC_k; \tag{15}$$

as may be established by induction. For from the foregoing, it follows that

$$(-1)^k\Delta^k(n)^{k+1} = {}_kC_0^n, -{}_kC_1^n, \cdots, (-1)^k{}_kC_k^n. \tag{16}$$

Then, since ${}_{k+1}C_r = {}_kC_r + {}_kC_{r-1}$, we conclude that

$$F_n^{k+1} = (-1)^{k+1}(n)^{k+1} = {}_{k+1}C_0^n, 0^{n-1}, -{}_{k+1}C_1^n, 0^{n-1}, \cdots, (-1)^{k+1}{}_{k+1}C_{k+1}^n. \tag{17}$$

If now $n \geqq 2$, then from (5) and (15) we find that

$$\cos\theta = 0; \qquad \theta/360° = 1/4. \tag{18}$$

Thus, the expected frequency of the changes in sign of $\Delta^k(n)^k$ is the same as that for the raw or ungraduated data. Moreover, if $n \geqq 3$, (8) leads to $\cos \theta' = -1/2$, found for the data For, in this case, at least two zero coefficients intervene between any two non-zero coefficients. And thus

$$\cos \theta' = -\sum_{r=0}^{k} {}_kC_r^2 \Big/ 2 \sum_{r=0}^{k} {}_kC_r^2 = -1/2. \tag{19}$$

In fact, the same factor cancels from numerator and denominator as we take higher differences, if a sufficient number of zeros intervene. More explicitly stated, the formula (9) found for the data is valid also for $\Delta^k(n)^k$, provided $n \geqq t + 2$.

To make this more concrete, it may be noted that cycle lengths corresponding to $t = 0, 1, 2, 3$, and 4, are respectively

$$4, 3, 2.73, 2\,60, 2.52. \tag{20}$$

From (15), we see directly that an element of $\Delta^k(n)^k$ is correlated only with certain other elements which are at distances from it which are multiples of n

Some of the foregoing results may be included in a theorem as follows: THEOREM: *Given a sequence of independent chance variates, each subject to the normal distribution* (3) *with mean zero. Upon this material, let k summings or averagings by n be performed and k differencings, in any order. Then the resulting sequence has something of the same chaotic nature as the data. In particular for* $n \geqq 2$ *the expected frequency of changes of sign is the same,—viz.,* 1/4 *for change from minus to plus and* 1/4 *for change from plus to minus. Moreover, as n is increased from* 2 *to* 3, 4, 5, $\cdots$, *the expected frequency of other characteristics becomes the same, maxima and minima, points of inflection, etc., in accordance with* (9).

But, suppose now that after $k + 1$ summings by n, only k differencings are performed. Is the resulting sequence almost chaotic? Hardly so. At least, it can be shown that changes of sign in each direction have no longer an expected frequency fixed at 1/4, but this expected frequency decreases as n increases. To show this, formula (5) is applied to (16); and setting in (10), $C = {}_{2k}C_k$, $C' = {}_{2k}C_{k-1}$ it follows that

$$\cos \theta = [(n - 1)C - C']/nC = 1 - (2k + 1)/n(k + 1). \tag{21}$$

Then $\cos \theta > 1 - 2/n$, and the cycle length for expected changes of sign in definite direction is somewhat greater than that obtained by setting $\cos \theta = 1 - 2/n$. For values of n not too small, we may write $\cos \theta = 1 - \theta^2/2$, approximately; and then approximately

$$\text{cycle length for definite change of sign in } \Delta^k(n)^{k+1} \text{ is } \pi\sqrt{n}. \tag{22}$$

If $n = 9$, this approximate length is 9.4, assuming k fairly large, whereas the more exact length is 9.2.

Consider now the result of summing $k + 2$ times, and then differencing only k

times. For this purpose, a few formulas for summing squares will be useful. By the method of differences it can be shown that if $l = a + nh$, and

$$T = a^2/2 + (a + h)^2 + (a + 2h)^2 + \cdots + (a + \overline{n - 1}h)^2 + l^2/2, \tag{23}$$

then

$$T = n(a^2 + al + l^2)/3 + (l - a)^2/6n. \tag{24}$$

Suppose, now, that a/n takes on the values $0, {}_kC_0, -{}_kC_1, \cdots, (-1)^k {}_kC_k$ in succession, while l/n takes on the values ${}_kC_0, -{}_kC_1, \cdots, (-1)^k {}_kC_k, 0$. Let U be the sum of the $(k + 1)$ values of T thus obtained. Then by (10),

$$U = n^3(2\,{}_{2k}C_k - {}_{2k}C_{k-1})/3 + n\sum_{i=0}^{k+1} {}_{k+1}C_i^2/6. \tag{25}$$

$$U = \frac{n^3}{3}\frac{(k + 2)(2k)!}{k!(k + 1)!} + \frac{n}{6}\,{}_{2k+2}C_{k+1}. \tag{26}$$

Now, by applying to (16) one more summation by n, there are formed $(k + 2)$ arithmetic progressions of $(n + 1)$ terms each, alternately increasing and decreasing. The maximum and minimum terms at the juncture of the progressions are to be split into two halves to apply (23). Then the sum of the squares of these coefficients is given by (26). This forms a denominator for (5).

To obtain the numerator for (5) we note that from $ab = [a^2 + b^2 - (a - b)^2]/2$ it follows that if

$$V = a(a + h) + (a + h)(a + 2h) + \cdots + (a + \overline{n - 1}h)(a + nh); \tag{27}$$

then, from (23),

$$V = T - nh^2/3 = T - (l - a)^2/3n. \tag{28}$$

If now W is the sum of such V's, reference to the last terms of (24) and (26) shows that

$$W = U - (n/3)\,{}_{2k+2}C_{k+1}. \tag{29}$$

And hence, from (5),

$$\cos\theta = \frac{(k + 2)n^2 - 4k - 2}{(k + 2)n^2 + 2k + 1}. \tag{30}$$

Then

$$\cos\theta > \frac{n^2 - 4}{n^2 + 2}; \tag{31}$$

but only slightly greater when k is large. Again

$$\cos\theta > 1 - 6/n^2; \tag{32}$$

but only slightly greater when n is not small. In this case, $\cos\theta = 1 - \theta^2/2$, approximately. And thus, approximately, for large k, and for n not small

$$\text{cycle length for definite change of sign of } \Delta^k(n)^{k+2} = 1.81n. \tag{33}$$

This gives for $n = 10$ a cycle length of 18.1; whereas, if $\cos\theta$ is taken as the right member of (31), the cycle length is 18.2.

Thus, if a $(k + 2)$-fold summation or averaging of random data is followed

THE CYCLIC EFFECTS OF LINEAR GRADUATIONS PERSISTING IN THE DIFFERENCES OF THE GRADUATED VALUES

By Edward L Dodd

University of Texas

1. Scope of inquiry. Slutzky [1] applied the moving sum, the repeated moving sum, and other linear processes to random numbers obtained from lottery drawings But the graph of the *moving sum* becomes, when the vertical scale is changed in the ratio of n to 1, the graph of the *moving average*, the simplest form of *graduation*. When cyclic effects are studied, there is no essential difference between a moving sum and a moving average, nor between a general linear process with coefficients $a_1, a_2, \cdots, a_s$, having sum $A \neq 0$ and the corresponding *graduation*, with coefficients $a_i' = a_i/A$. Thus Slutzky's work throws considerable light upon graduation, although his main interest was in summation.

Slutzky found that the graphs of moving sums of random numbers bore strong resemblance to graphs of economic phenomena, such as [1, p. 110] that of English business cycles from 1855 to 1877. In fact, Slutzky regards the fluctuations in economic phenomena as due largely to a synthesizing of random causes.

In general the undulatory character of such values cannot be described as periodic; since the waves are of different length. But Slutzky found that, upon operating on random data having mean zero and constant variance, the resulting values approach a sinusoidal limit under certain conditions,—in particular, when a set of n summations by twos is followed by m differencings, and as $n \to \infty$, $m/n \to a$ constant Romanovsky [2] generalized this result by taking successive summations of s consecutive elements of the data, with $s \geqq 2$; but required that $m/n \to \alpha \neq 1$. However, the cases which are of interest to me just now are those for which $m = n - 1$ or $m = n - 2$; and for these cases $m/n \to 1$. Romanovsky considers the case of $m = n - 1$,—not, however, as leading to a sinusoidal limit,—and gives in formula (46) the value of a coefficient of correlation—which I deduce directly From his formula (43) a corresponding coefficient of correlation can be obtained for the case of $m = n - 2$, as the sum of certain products. A more simple expression than this I need, which I obtain directly. In my treatment, these coefficients are the cosines of angles; and the ratio of such an angle to a whole revolution is an expected frequency of occurrence.

After setting forth in Section 2 some preliminary formulas, I treat in Section 3 the results of applying to random data an indefinite number $k + 2$ of summations or averagings, followed by k differencings—the number of terms in a sum remaining fixed. In Section 4, however, only a few differencings are applied to a

graduation. In particular the Spencer 21-term formula is studied in some detail. In former papers [3, 4] I have dealt with the immediate effects of graduations upon random data.

The question to be considered in this paper is this: *Do the cyclic effects appearing in the graduated values persist in the successive differences? And, if so, do these affects fade out gradually or on the other hand, do they come to a rather abrupt termination?*

These differences of graduated values, indeed, up to the third, fourth or fifth are of considerable importance. Henderson [5] defines the smoothing coefficient of a given graduation as the ratio of the theoretical standard deviation of the third differences for the graduated values to that for the original values or data.

2. Preliminary notions and formulas. The data to be graduated will be supposed to be independent, or uncorrelated, or as Slutzky expresses it, "incoherent." This will imply that the expected value of the product of two different chance variates is the product of their expected values.

Now the operations of summing and differencing as used here are not inverse. To illustrate: Given as independent $u, v, w, x, y, z, \cdots$. Summing by twos yields the sequence $u + v, v + w, w + x, x + y, y + z, \cdots$. But the first differences of these numbers, $w - u, x - v, y - w, z - x, \cdots$ are alternately correlated, thus $w - u$ is negatively correlated with $y - w$; $x - v$ with $z - x$, etc. Indeed, successive differencing following successive summing does not lead back to the original condition of incoherency. However, under certain conditions, the resulting coherency may be so slight that the final succession of numbers may have just about the same chaotic properties as the succession of data.

In my paper [3, p. 262], I set forth a number of features on the basis of which a cycle length could be defined. One of these involves the frequency of maxima. Given independent chance variables, each subject to the same law of distribution,

$$P(x_j \leqq x) = \Phi(x); \tag{1}$$

where $\Phi(x)$ has a derivative $\phi(x)$. It is then easy to see that the expected relative frequency of maxima is 1/3. That is:

$$P(x_{i-1} \leqq x_i \geqq x_{i+1}) = \int_{-\infty}^{\infty} [\Phi(x)]^2 \phi(x)\, dx = 1/3. \tag{2}$$

Now, for a given feature, a cycle length is defined as the reciprocal of the theoretic relative frequency. Then the cycle length here for *maxima* is three. It is well known that averaging tends to remove maxima. Thus, upon averaging or summing, the cycle length tends to increase. It is almost as well known that differencing tends to increase the frequency of maxima, and thus decrease cycle length. For if $z_i = \Delta y_i = y_{i+1} - y_i$, then between two maxima of y_i, there is at least one minimum (strong and weak) of y_i; and following this minimum and before passing the next maximum of y_i there is *at least one* maximum of z_i. Successive differencing tends to reduce the cycle length of maxima from 3 to 2,

by only k differencings, the resulting graduation or linear processing $z = \Delta^k(n)^{k+2}$ is decidedly not as chaotic as the data; as seen from (31) and (33). But further, $\Delta z = \Delta^{k+1}(n)^{k+2}$; and thus from (22) the cycle length for the expected maxima of z is about $\pi\sqrt{n}$.

Now Slutzky [1, p. 109] distinguished conspicuous waves from inconsequential "ripples." On this basis, the frequency of significant cyclical features for a chance variable, such as z, would be less than the frequency of the maxima. It is not so clear that the frequency of significant features of a chance variable will be *greater* than that for changes of sign in definite direction. That turned out to be true for graduated values such as discussed in my earlier paper [3, p. 262]. If this be also valid for z, we would expect that conspicuous "waves" of $\Delta^k(n)^{k+2}$ would have average length between $\pi\sqrt{n}$ and 1 81n, except for small values of n and k

4. Graduations or linear processes and their successive differences. If double summation by n is followed by a single differencing, the result—as indicated in (14)—is, for $n = 3$,

$$y_i = -x_i - x_{i+1} - x_{i+2} + x_{i+3} + x_{i+4} + x_{i+5}. \tag{34}$$

Then

$$y_{i+3} = -x_{i+3} - x_{i+4} - x_{i+5} + x_{i+6} + x_{i+7} + x_{i+8}. \tag{35}$$

Thus y_i and y_{i+3} are negatively correlated; since x_{i+3}, x_{i+4}, and x_{i+5} appear in each, but with sign changed. This would seem to tend to make maxima alternate with minima at distances of about 3; or at distances of n, in the general case (14). Here, following Slutzky and Romanovsky, the coefficient of correlation r_p between elements at a distance of p is taken as

$$r_p = E(x_r \cdot x_{r+p})/E(x_r)^2. \tag{36}$$

Using computed averages, instead of expected values, Alter [6] recommends a "correlation periodogram," in which r_p is the ordinate for abscissa p.

Moreover, we would expect a graduation (4) with coefficients a_i proportional to the ordinates y of the sinusoid $y = \sin(\alpha + 2\pi x/p)$ taken for $x = 1, 2, 3, \cdots$ to impress upon random data oscillations with maxima separated from minima by about $p/2$. But such a_i, as well as those in (34), have abrupt endings which introduce noticeable alterations. More satisfactory results come from tapering ends, such as appear in damped vibration, with coefficients about proportional to $e^{-\alpha|x|}\cos 2\pi x/p$ or to $e^{-\alpha|x|}\sin 2\pi x/p$. H. Labrouste and Mrs. Labrouste [7] give a powerful operator of this description.

Slutzky (loc. cit. pp. 119–123), Yule [8], and Walker [9] make use of damped harmonic vibration to explain the creation of cycles; while Bartels [10] approaches by a different method the oscillations that do not last.

Now the common graduation formulas have coefficients not conforming strictly to damped vibration, as the tapering ends vibrate more quickly. However, these ends have little more than a smoothing or stabilizing effect. Furthermore,

the coefficients for first differences are likely to conform to something like $e^{-c|x|} \sin 2\pi x/p$. Some *experimental* evidence will be presented for the following conclusion:

If the coefficients a_i of a graduation or linear process (4) appear to conform roughly to equidistant ordinates of a damped vibration, $\pm e^{-c|x|} \cos 2\pi x/p$ or $\pm e^{-c|x|} \sin 2\pi x/p$, with changes of sign at intervals of $p/2$, then when this process (4) is applied to independent chance data having zero mean and constant variance, there is a tendency for the graduated or processed values to change sign at intervals of about $p/2$.

A number of standard graduations have first and second differences see (6), (7)—which bear a decided resemblance to damped vibrations, while the third or fourth differences have only moderate, if any, cyclic appearance. This is especially true of those graduations which are constructed by applying three summings—the number of terms in a sum being in general different and a fourth

TABLE I

Coefficients (×350) for Spencer 21-term graduation and for first four differences. Also theoretical cycle lengths for change in sign in values obtained from random data

		Cycle Length
Grad.	$+$ 6,18,33,47,57,60,57,47,33,18,6 / $-$ 1,3,5,5,2 2,5,5,3,1	10.7
1st D.	$+$ 1,2,2,0 3,10,14,15,12,8,3 / $-$ 3,8,12,15,14,10,3 0,2,2,1	7.0
2nd D.	$+$ 2,3,5,4,3 3,4,5,3,2 / $-$ 1,1,0 1,4,7,6,7,4,1 0,1,1	5.5
3rd D.	$+$ 1,0 1,1,4,3,3 1 2,1,2,1 / $-$ 1,2,1,2, 1 3,3,4,1,1 0,1	3.2
4th D.	$+$ 1,1,1 1 0 1 4 4 1 0 1 1,1,1 / $-$ 1 1 3 3 0 2 0 3 3 1 1	1.6

process with negative coefficients. This is, indeed, a favorite form of graduation, with which are associated the names of Woolhouse, Spencer, Higham, Kenchington, Henderson, etc. The Spencer 21-term formula, for which some features have already been described, [3, p. 262], will now be examined, with special reference to its differences. Cycle length for change of sign is one-half that for change from minus to plus.

In the graduation formula, itself, there are 11 positive coefficients, centrally located, and relatively large as compared with the negative coefficients. This 11 is close to 10.7 the theoretical cycle length for changes of sign of $y_r - 4.5$, the difference between the graduated value y_r and its mean—the arithmetic mean of $1, 2, \cdots, 9$. The structure of the first and second differences also

matches closely the corresponding cycle lengths. In the third differences, there is a break at the center; but still there appears considerable regularity. But among fourth differences, the zigzag is the prominent feature. Now the theorem of Section 3 does not really apply to the Spencer formula, with its two summations by fives and one summation by sevens, and another process But it is not surprising that the cyclicity ceases after passing the third differences.

As a basis for comparing observed values with expected values, the tenth digits in the 600 logarithms from log 200 to log 799 were taken as a random set of numbers. These 600 numbers had been given a Spencer 21-term graduation [3, pp 261–262], yielding 580 graduated values. From these the 579 first differences were found, the 578 second differences, etc These numbers, 580, 579, $\cdots$, were multiplied respectively by the expected relative frequences of change in sign of $y_r - 4.5$, of Δy_r, $\Delta^2 y_r$, etc, as found by use of (5), (8), and similar expressions to form the following table.

The most abrupt change in frequency or cycle length appears to occur in passing from third to fourth differences. In Table I, this is seen in the configuration of positive and negative terms, and in the drop from 3.2 to 1.6 in cycle length; and in Table II in the corresponding increase in expected sign changes from 90.4 to 176.7. More spectacular is the increase in the number of zigzags represented by $-, +, -, +$. Among the third differences, there were found only 13 instances of four successive terms with signs as just indicated, whereas among fourth differences there were found 75 such instances. For random material, about 36 such zigzags would be expected—decidedly more than found among the third difference, and decidedly less than found among the fourth differences.

TABLE II

Comparison of expected changes of sign with observed changes for a Spencer 21-term graduation

	Expected Number of Changes from − to +	*Observed Number of Changes from − to +*
Graduated values—4.5	27.2	27
First differences	41.3	42
Second differences	52.9	48
Third differences	90.4	74
Fourth differences	176.7	146

The Spencer 21-term graduation appears to be fairly representative of commonly used graduations as regards regularity or irregularity in the distribution of positive and negative coefficients among the differences. For graduations with a much larger number of terms, the alternation of sign in fourth differences may not be so rapid, as, e g. in the 35-term 5th degree parabolic graduation which Macaulay [11] calls No. 18 On the other hand, for a formula with non-tapering ends, such as the 13-term formula which Macaulay gives [11,

p. 64], the coefficients appearing in the differences are more irregular, especially at the ends. While the Spencer formula is fairly representative, different formulas have distinguishing features. If it is desirable to form an idea of what a given formula will do to random data, a table like Table I can be constructed.

5. Summary. When upon independent chance data, summing, averaging or some more general graduation process is used, the graduated values tend to assume a wavy configuration. These waves often seem to have a fair amount of regularity or cyclicity. The first differences usually, and often other differences of the graduated values, are decidedly cyclic But, as we go in turn to the higher differences, the cyclicity may weaken. Indeed there may be a return to something like randomness. And subsequent differencings may tend to set up zigzags.

If $(k + 2)$ successive summings by n have been performed on independent chance data, with n not too small, say $n \geqq 5$—then $k + 2$ differencings will just about bring back the original chaotic or random condition. But with only k or $(k + 1)$ differencings, a definite cyclicity remains, at least theoretically, in the expected values.

In the case of the Spencer 21-term graduation, the coefficients for the successive differences indicate the appearance of cyclicity in first, second, and third differences.

REFERENCES

[1] Eugen Slutzky, "The summation of random causes as the source of cyclic processes," *Econometrica*, Vol. 5 (1937), pp. 105-146. This supplements an earlier paper (1927) in Russian.

[2] V. Romanovsky, "Sur la loi sinusoidale limite," *Rendiconti del Circolo Matematico di Palermo*, Vol. 56 (1932), pp 82-111.

[3] Edward L Dodd, "The length of the cycles which result from the graduation of chance elements," *Annals of Math Stat.*, Vol. 10 (1939), pp. 254-264.

[4] Edward L. Dodd, "The problem of assigning a length to the cycle to be found in a simple moving average and in a double moving average of chance data," *Econometrica*, Vol 9 (1941), pp. 25-37.

[5] Robert Henderson, *Graduation of Mortality and Other Tables*, Actuarial Society of America, New York, 1919.

[6] Dinsmore Alter, "A group or correlation periodogram, with applications to the rainfall of the British Isles," *Monthly Weather Review*, Vol. 55 (1927), pp. 263-266.

[7] H. and Mrs. Labrouste, "Harmonic analysis by means of linear combinations of ordinates," *Terr. Mag. and Atmos. Elec.*, Vol. 41 (1936), pp. 17-18.

[8] G. Udny Yule, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Phil Trans* A, Vol. 226 (1927), pp. 267-298.

[9] Sir Gilbert Walker, "On periodicity in series of related terms," *Roy. Soc. Proc.*, Ser. A, Vol. 131 (1931), pp. 518-532.

[10] J Bartels, "Random fluctuations, persistence, and quasi-persistence in geophysical and cosmical periodicities," *Terr. Mag. and Atmos. Elec.*, Vol 40 (1935), pp. 1-60.

[11] F R Macaulay, *The Smoothing of Time Series*, Publication of the National Bureau of Economic Research, No 19, New York, 1931.

ON THE DISTRIBUTION OF WILKS' STATISTIC FOR TESTING THE INDEPENDENCE OF SEVERAL GROUPS OF VARIATES

By A. Wald[1] and R. J. Brookner[1]

Columbia University

1. **Introduction.** We consider p variates $x_1, x_2, \cdots, x_p$ which have a joint normal distribution Let the variates be divided into k groups; group one containing $x_1, x_2, \cdots, x_{p_1}$, group two containing $x_{p_1+1}, x_{p_1+2}, \cdots, x_{p_2}$, etc. We are interested in testing the hypothesis that the set of all population correlation coefficients between any two variates which belong to different groups is zero.

Wilks[2] has derived, by using the Neyman-Pearson likelihood ratio criterion, a statistic based on N independent observations on each variate with which one may test this hypothesis Let $||r_{ij}||$ be the matrix of sample correlation coefficients; Wilks' statistic, λ, is the ratio of the determinant of the p-rowed matrix of sample correlations to the product of the p_1-rowed determinant of correlations of the variates of group one, the $(p_2 - p_1)$-rowed determinant of correlations of the second group, etc That is

$$\lambda = \frac{|r_{ij}|}{|r_{\alpha_1\beta_1}| \cdot |r_{\alpha_2\beta_2}| \cdots |r_{\alpha_k\beta_k}|}$$

where $|r_{\alpha_i\beta_i}|$ is the principal minor of $|r_{ij}|$ corresponding to the ith group.

In order to use the test, the distribution function of λ must be known. Wilks has shown that in certain cases the exact distribution is a simple elementary function; in other cases it is an elementary function, but one which is rather unwieldy and which does not lend itself readily to practical use. It is our purpose in this paper (1) to show a method by which the exact distribution can be explicitly given as an elementary function for a certain class of groupings of the variates, and (2) to give an expansion of the exact cumulative distribution function in an infinite series which is applicable to any grouping.

2. **The exact distribution of λ.** By the method to be described, the exact distribution of λ can be found when the numbers of variates in the groups are such that there are an odd number in at most one group. If the number of variates is small, say at most eight, the method will increase only slightly the list of distribution functions that Wilks gives in his paper.

---

[1] Research under a grant-in-aid of the Carnegie Corporation of New York.

[2] S. S. Wilks, "On the independence of k sets of normally distributed statistical variables," *Econometrica*, Vol. 3 (1935), pp 309-326. Other references to Wilks in this paper except where otherwise noted are to this publication

For purposes of deriving the distribution of λ we may assume that $E(x_u) = 0$, $(u = 1, 2, \cdots, p)$; that there are $n = N - 1$ independent observations $x_{u\alpha}$ $(\alpha = 1, 2, \cdots, n)$ on each variate x_u; and that the sample covariance between x_i and x_j is given by $s_{ij} = \sum_{\alpha=1}^{n} x_{i\alpha}x_{j\alpha}/n$. We define u' (a function of u) to be the total number of variables in all the groups which precede the group in which x_u lies. The complete theory is independent of the ordering of the groups and of the ordering of the variates within the groups; hence without loss of generality, we may assume that if any group contains an odd number of variates, it will be the last group, hence u' is always an even integer.

Wilks has shown that λ is a product $\prod_{u=p_i+1}^{p} z_u$ where each z_u is distributed independently of the others, and that the distribution of z_u is

$$\frac{z_u^{\frac{1}{2}(n-u-1)}(1-z_u)^{\frac{1}{2}(u'-2)}}{\mathrm{B}[\frac{1}{2}(n-u+1),\ u'/2]}\,dz_u \tag{1}$$

Now let $y_u = \log z_u$, then the characteristic function of y_u is

$$\begin{aligned}\phi_u(t) &= \frac{1}{\mathrm{B}[\frac{1}{2}(n-u+1),\ u'/2]}\int_0^1 e^{t\log z_u} z_u^{\frac{1}{2}(n-u-1)}(1-z_u)^{\frac{1}{2}(u'-2)}\,dz_u \\ &= \frac{1}{\mathrm{B}[\frac{1}{2}(n-u+1),\ u'/2]}\int_0^1 z_u^{\frac{1}{2}(n-u-1)+t}(1-z_u)^{\frac{1}{2}(u'-2)}\,dz_u\end{aligned}$$

where t is a pure imaginary. It is known[3] that this integral, even with complex exponents, is the Beta-function so long as the real parts of both exponents are greater than minus one, so

$$\begin{aligned}\phi_u(t) &= \frac{\mathrm{B}[\frac{1}{2}(n-u+1)+t,\ u'/2]}{\mathrm{B}[\frac{1}{2}(n-u+1),\ u'/2]} \\ &= \frac{\Gamma[\frac{1}{2}(n-u+1)+t]\,\Gamma[\frac{1}{2}(n-u+1+u')]}{\Gamma[\frac{1}{2}(n-u+1+u')+t]\,\Gamma[\frac{1}{2}(n-u+1)]}.\end{aligned} \tag{2}$$

But here u' is always an even integer, hence by the well known recursion formula of the Gamma-function, which is valid for complex arguments excluding only negative integers

$$\phi_u(t) = c_u\{[\tfrac{1}{2}(n-u+1)+t][\tfrac{1}{2}(n-u+3)+t] \cdots [\tfrac{1}{2}(n-u+u'-1)+t]\}^{-1}$$

where

$$c_u = [\tfrac{1}{2}(n-u+1)][\tfrac{1}{2}(n-u+3)] \cdots [\tfrac{1}{2}(n-u+u'-1)].$$

[3] See Whittaker and Watson, *A Course in Modern Analysis*, Fourth edition 1927, Chap. 12.

Now set

$$y = \log \lambda = y_{p_1+1} + y_{p_1+2} + \cdots + y_p$$

and the characteristic function of y is

$$\phi(t) = \prod_{u=p_1+1}^{p} c_u\{[\tfrac{1}{2}(n-u+1)+t][\tfrac{1}{2}(n-u+3)+t] \cdots [\tfrac{1}{2}(n-u+u'-1)+t]\}^{-1}.$$

From the characteristic function, we can obtain the distribution function, $g(y)$, of y by the relation

$$g(y) = \frac{c_n}{2\pi i}\int_{-i\infty}^{i\infty} \frac{e^{-yt}\,dt}{\prod_{u=p_1+1}^{p}[\tfrac{1}{2}(n-u+1)+t]\cdots[\tfrac{1}{2}(n-u+u'-1)+t]} = \frac{c_n}{2\pi i}\int_{-i\infty}^{i\infty}\Phi(t)\,dt,$$

where

$$c_n = \prod_{u=p_1+1}^{p} c_u.$$

The integration can be carried out by the method of residues; since y is always negative (the range of λ is from 0 to 1), on a half circle with center at the origin in the negative half of the complex t-plane, the integral of the function $\Phi(t)$ converges to zero as the radius of the circle becomes infinite. Since $\Phi(t)$ is analytic except for a finite number of poles on the negative real axis, $g(y)$ is c_n times the sum of the residues at these points.

Now $\Phi(t)$ is of the form $\dfrac{e^{-yt}}{P(t)}$ where $P(t)$ is a polynomial in t as follows: suppose that the groups contain $r_1, r_2, \cdots, r_k$ variables respectively, then let (k_j+1) be the number of these r's which are greater than or equal to j; then

$$P(t) = [\tfrac{1}{2}(n-2)+t]^{k_1}[\tfrac{1}{2}(n-3)+t]^{k_2}[\tfrac{1}{2}(n-4)+t]^{k_3+k_1}[\tfrac{1}{2}(n-5)+t]^{k_4+k_2}$$
$$[\tfrac{1}{2}(n-6)+t]^{k_5+k_3+k_1}\cdots[\tfrac{1}{2}(n-p+1)+t]^{k_{p-2}+k_{p-4}+\cdots+k_{[\frac{1}{2}p]-[\frac{1}{2}(p-3)]}},$$

where

$$[\sigma/2] = \begin{cases} \sigma/2 \text{ if } \sigma \text{ is even} \\ (\sigma-1)/2 \text{ if } \sigma \text{ is odd.} \end{cases}$$

Then

$$g(y; r_1, r_2, \cdots, r_k) = c_n \sum_{\alpha=1}^{p-2}\frac{1}{\theta_\alpha!}\frac{d^{\theta_\alpha}}{dt^{\theta_\alpha}}[(t+\tfrac{1}{2}(n-\alpha-1))^{\theta_\alpha+1}\Phi(t)]_{t=-\frac{1}{2}(n-\alpha-1)}$$

where

$$\theta_\alpha + 1 = k_\alpha + k_{\alpha-2} + \cdots + k_{[\frac{1}{2}(\alpha+2)]-[\frac{1}{2}(\alpha-1)]}.$$

It can be shown that θ_α is $\geqq 0$ for α between 1 and $p-2$. Thus we have $g(y; r_1, r_2, \cdots, r_l)$ and from it we can calculate $f(\lambda; r_1, r_2, \cdots, r_l)$.

Suppose $p = 8$ and that the variables are divided into two groups of four each, then we will calculate the distribution function $f(\lambda; 4, 4)$. Now

$$g(y; 4, 4) = \frac{c_n}{2\pi i}\int_{-\infty}^{+\infty} \frac{e^{-yt}\,dt}{[\frac{1}{2}(n-2)+t][\frac{1}{2}(n-3)+t][\frac{1}{2}(n-4)+t]^2 \cdot [\frac{1}{2}(n-5)+t]^2[\frac{1}{2}(n-6)+t][\frac{1}{2}(n-7)+t]}$$

and

$$c_n = \left(\frac{n-2}{2}\right)\left(\frac{n-3}{2}\right)\left(\frac{n-4}{2}\right)^2\left(\frac{n-5}{2}\right)^2\left(\frac{n-6}{2}\right)\left(\frac{n-7}{2}\right).$$

Then

$$g(y; 4, 4) = 16c_n\left[\frac{-e^{\frac{1}{2}(n-2)y}}{90} + e^{\frac{1}{2}(n-3)y} + \frac{8e^{\frac{1}{2}(n-4)y}}{9} - \frac{8e^{\frac{1}{2}(n-5)y}}{9} - e^{\frac{1}{2}(n-6)y} + \frac{e^{\frac{1}{2}(n-7)y}}{90} - \frac{ye^{\frac{1}{2}(n-4)y}}{3} + \frac{ye^{\frac{1}{2}(n-5)y}}{3}\right].$$

Since

$$y = \log\lambda, \qquad dy = \frac{d\lambda}{\lambda},$$

we have

$$f(\lambda; 4, 4,) = \frac{16c_n}{3}\left[-\frac{\lambda^{\frac{1}{2}(n-4)}}{30} + \frac{\lambda^{\frac{1}{2}(n-5)}}{2} - \frac{8\lambda^{\frac{1}{2}(n-6)}}{3} + \frac{8\lambda^{\frac{1}{2}(n-7)}}{3} - \frac{\lambda^{\frac{1}{2}(n-8)}}{2} + \frac{\lambda^{\frac{1}{2}(n-9)}}{30} - (\lambda^{\frac{1}{2}(n-7)} + \lambda^{\frac{1}{2}(n-6)})\log\lambda\right].$$

The cumulative distribution function is given by

$$J_w(4, 4) = \text{Prob}\,[\lambda \leq w; 4, 4]$$
$$= \frac{16c_n}{3}w^{\frac{1}{2}(n-7)}\left[\frac{1}{15(n-7)} - \frac{w^{\frac{1}{2}}}{n-6} - \frac{4(4n-23)w}{3(n-5)^2} + \frac{14(4n-13)w^{\frac{1}{2}}}{3(n-4)^2} + \frac{w^2}{n-3} - \frac{w^{\frac{1}{2}}}{15(n-2)} - \left(\frac{2w}{n-5} + \frac{2w^{\frac{1}{2}}}{n-4}\right)\log w\right].$$

Wilks' expression for the cumulative distribution function appears to be quite different, but if we substitute $n = N-1$ and use the relation

$$\beta_{\sqrt{w}}(N-6; 4) = \frac{\Gamma(N-2)}{\Gamma(N-6)\cdot\Gamma(4)}\int_0^{\sqrt{w}} x^{N-7}(1-x)^3\,dx$$
$$= \tfrac{1}{6}(n-2)(n-3)(n-4)(n-5) \cdot\left[\frac{w^{\frac{1}{2}(n-5)}}{n-5} - \frac{3w^{\frac{1}{2}(n-4)}}{n-4} + \frac{3w^{\frac{1}{2}(n-3)}}{n-3} - \frac{w^{\frac{1}{2}(n-2)}}{n-2}\right]$$

it can be shown that the two formulas for the cumulative distribution are identical.

In cases where u' is not always an even integer, the exact distribution function of λ can still be obtained using this method. However, in such a case, the gamma functions do not cancel out and the integrand has an infinitude of poles, so the function is expressed by an infinite series. We will use a different method to obtain an infinite series expansion.

3. **A series expansion of the cumulative distribution function.** Let us put $v = -y$, and let the density function of v be $h(v)$, then from (2), we have

$$h(v)\,dv = dv\,\frac{c_n}{2\pi i}\int_{-i\infty}^{i\infty} e^{vt} \prod_{u=r_1+1}^{p} \frac{\Gamma[\frac{1}{2}(n-u+1)+t]\,dt}{\Gamma[\frac{1}{2}(n-u+1+u')+t]}.$$

Since v is a monotonic decreasing function of λ, and since the critical region for testing the null hypothesis is given by the inequality $\lambda < \lambda_0$, then the critical region will be defined by $v > v_0$, where v_0 is such that

$$\int_{v_0}^{\infty} h(v)\,dv$$

is equal to a chosen level of significance.

PROPOSITION 1.

$$h(v) = h_n(v)\bar{\psi}(v)$$

where $\bar{\psi}(v)$ does not depend on n, and $h_n(v) = c_n e^{-\frac{1}{2}v}$.

PROOF: Let

$$t' = t + \tfrac{1}{2}(n-p).$$

Then

$$h(v) = \frac{c_n}{2\pi i}\int_{-i\infty+\frac{1}{2}(n-p)}^{i\infty+\frac{1}{2}(n-p)} e^{v(t'-\frac{1}{2}(n-p))} \prod_u \frac{\Gamma[\frac{1}{2}(p-u+1)+t']\,dt'}{\Gamma[\frac{1}{2}(p-u+u'+1)+t']}.$$

Now the area in the complex plane bounded by the vertical line through $\frac{1}{2}(n-p)$, by the vertical line through the origin, and by arcs of a circle with center at the origin of arbitrary radius is one in which the integrand is everywhere regular. Furthermore, the integral along the arcs approaches zero as the radius of the circle approaches infinity, hence the integrals along the vertical line through $\frac{1}{2}(n-p)$ and along the vertical axis are equal. Then we may write

$$\begin{aligned}\frac{e^{\frac{1}{2}v}}{c_n}h(v) &= \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} e^{v(t'+p/2)} \prod_u \frac{\Gamma[\frac{1}{2}(p-u+1)+t']\,dt'}{\Gamma[\frac{1}{2}(p-u+u'+1)+t']} \\ &= \bar{\psi}(v).\end{aligned}$$

Therefore

$$h(v) = c_n e^{-\frac{1}{2}v}\bar{\psi}(v).$$

PROPOSITION 2.

$$I = \lim_{n\to\infty} \int_0^\infty \frac{c_n e^{-\frac{1}{2}v} v^{r-1}\, dv}{\Gamma(r)} = 1$$

where we define

$$r = \sum_{j=k+1}^{\lambda} \sum_{i=1}^{k-1} \frac{r_i r_j}{2}$$

so that

$$r = \tfrac{1}{2}[r_2 r_1 + r_3(r_1 + r_2) + \cdots + r_k(r_1 + r_2 + \cdots + r_{\lambda-1})]$$
$$= \tfrac{1}{2} \sum_u u'.$$

PROOF: Let

$$\frac{n}{2} v = v^*$$

then

$$\int_0^\infty c_n e^{-\frac{1}{2}v} v^{r-1}\, dv = \int_0^\infty c_n e^{-v^*} \left(\frac{2}{n}\right)^r (v^*)^{r-1}\, dv^*$$
$$= c_n \left(\frac{2}{n}\right)^r \Gamma(r).$$

Hence

$$I = \lim_{n\to\infty} c_n \left(\frac{2}{n}\right)^r$$

but

$$c_n = \prod_u \frac{\Gamma\frac{1}{2}(n - u + 1 + u')}{\Gamma\frac{1}{2}(n - u + 1)}$$

and therefore

$$I_u = \lim_{n\to\infty} \frac{\Gamma\frac{1}{2}(n - u + 1 + u')}{\Gamma\frac{1}{2}(n - u + 1)} \left(\frac{2}{n}\right)^{u'/2} = 1$$

by an application of the Stirling approximation. Therefore

$$I = \prod_u I_u = 1.$$

We then write

$$\psi(v) = \frac{\bar{\psi}(v)\Gamma(r)}{v^{r-1}}$$

hence

$$h(v) = \frac{c_n e^{-\frac{n}{2}v} v^{r-1} \psi(v)}{\Gamma(r)}. \tag{3}$$

PROPOSITION 3. *For any positive integer s,*

$$\lim_{n\to\infty} \left\{ n^s \cdot \text{Prob}\left(v > \frac{1}{\sqrt{n}} \right) \right\} = 0.$$

PROOF: Since $v = -\log \lambda$, the inequality $v > 1/\sqrt{n}$ is equivalent to the inequality $\lambda < e^{-1/\sqrt{n}}$. Since $\lambda = \prod_{u=p_1+1}^{p} z_u$, the inequality $\lambda < e^{-1/\sqrt{n}}$ implies that there exists at least one value of u for which

$$z_u < e^{-1/(p-p_1)\sqrt{n}}.$$

Hence

$$\sum_{u=p_1+1}^{p} P(z_u < e^{-1/(p-p_1)\sqrt{n}}) \geq P(\lambda < e^{-1/\sqrt{n}}) = P(v > 1/\sqrt{n}).$$

Hence in order to prove Proposition 3 we have only to show that for each u and any arbitrary positive integer s

$$\lim_{n\to\infty} \{ n^s \cdot P(z_u < e^{-1/(p-p_1)\sqrt{n}}) \} = 0.$$

From (1) we have

$$P(z_u < e^{-1/(p-p_1)\sqrt{n}})$$

$$= \frac{1}{\text{B}[\frac{1}{2}(n-u+1); u'/2]} \int_0^{e^{-1/(p-p_1)\sqrt{n}}} z_u^{\frac{1}{2}(n-u-1)} (1-z_u)^{\frac{1}{2}(u'-2)} \, dz_u.$$

Over the range of integration, we have $z_u \leq e^{-1/(p-p_1)\sqrt{n}}$ so

$$P(z_u < e^{-1/(p-p_1)\sqrt{n}}) \leq \frac{e^{\frac{1}{2}(n-u-1)/(p-p_1)\sqrt{n}}}{\text{B}[\frac{1}{2}(n-u+1); u'/2]} \int_0^{e^{-1/(p-p_1)\sqrt{n}}} (1-z_u)^{\frac{1}{2}(u'-2)} \, dz_u$$

$$= \frac{e^{-\frac{1}{2}(n-u-1)/(p-p_1)\sqrt{n}}}{\text{B}[\frac{1}{2}(n-u+1); u'/2]} \left[-\frac{2}{u'} (1-z_u)^{u'/2} \right]_0^{e^{-1/(p-p_1)\sqrt{n}}}$$

$$= \frac{2e^{-\frac{1}{2}(n-u-1)/(p-p_1)\sqrt{n}}}{u' \cdot \text{B}[\frac{1}{2}(n-u+1); u'/2]} [1 - (1 - e^{-1/(p-p_1)\sqrt{n}})^{u'/2}].$$

It follows from the Stirling formula that

$$\lim_{n\to\infty} \left(\frac{n}{2}\right)^{u'/2} \text{B}[\tfrac{1}{2}(n-u+1); u'/2] = \lim_{n\to\infty} \frac{\Gamma\frac{1}{2}(n-u+1)\Gamma(u'/2)}{\Gamma\frac{1}{2}(n-u+u'+1)} \left(\frac{n}{2}\right)^{u'/2}$$

$$= \Gamma(u'/2).$$

Since

$$\lim_{n\to\infty} n^{\frac{1}{2}u'+s} e^{-\sqrt{n}/2(p-p_1)} = 0$$

and

$$\lim_{n\to\infty} (1 - (1 - e^{-1/\sqrt{n}})) = 1,$$

the proposition follows.

PROPOSITION 4. *The function $\psi(v)$ of formula (3) can be expanded in a power series, i.e.*

$$\psi(v) = \alpha_0 + \alpha_1 v + \alpha_2 v^2 + \cdots$$

with a finite radius of convergence.

PROOF: Wilks[4] has considered the following integral equation:

$$\int_0^B w^t g(w)\, dw = CB^t \frac{\Gamma(b_1 + t)\cdot\Gamma(b_2 + t) \cdots \Gamma(b_q + t)}{\Gamma(c_1 + t)\cdot\Gamma(c_2 + t) \cdots \Gamma(c_q + t)},$$

where $C = \dfrac{\Gamma(c_1)\cdot\Gamma(c_2) \cdots \Gamma(c_q)}{\Gamma(b_1)\cdot\Gamma(b_2) \cdots \Gamma(b_q)}$, B and $g(w)$ are independent of t, and $b_i < c_i$ $(i = 1, 2, \cdots, q)$. Wilks has shown that the solution of the integral equation, $g(w)$, is given by the following expression:

$$\begin{aligned} g(w) = {} & \frac{kw^{b_q-1}\left(1 - \dfrac{w}{B}\right)^{\gamma_q-\beta_q-1}}{B^{b_q}} \int_0^1 \int_0^1 \cdots \int_0^1 v_1^{c_1-b_1-1} v_2^{c_2-b_2-1} \cdots v_{q-1}^{c_{q-1}-b_{q-1}-1} \\ & \times (1 - v_1)^{\gamma_q-\beta_{q-1}-1}(1 - v_2)^{\gamma_{q-1}-\beta_{q-2}-1} \cdots (1 - v_{q-1})^{\gamma_1-\beta_1-1} \\ & \times \left[1 - v_1\left(1 - \frac{w}{B}\right)\right]^{b_1-c_2} \left[1 - \{v_1 + v_2(1 - v_1)\}\left(1 - \frac{w}{B}\right)\right]^{b_2-c_3} \cdots \\ & \times \left[1 - \{v_1 + v_2(1 - v_1) + \cdots \right. \\ & \qquad \left. + v_{q-1}(1 - v_1)(1 - v_2) \cdots (1 - v_{q-2})\}\left(1 - \frac{w}{B}\right)\right]^{b_{q-1}-c_q} \\ & \times dv_1\, dv_2 \cdots dv_{q-1} \end{aligned} \tag{4}$$

where

$$k = \prod_{i=1}^{q} \frac{\Gamma(c_i)}{\Gamma(b_i)\Gamma(c_i - b_i)}$$

and

$$\gamma_i = \sum_{j=0}^{i-1} c_{q-j} \qquad \beta_i = \sum_{j=0}^{i-1} b_{q-j}$$

[4] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol 24 (1932), pp 474-5.

the range of w being $0 \leqq w \leqq B$. Wilks has furthermore shown that

$$(5) \quad \{v_1 + v_2(1 - v_1) + \cdots + v_i(1 - v_1)(1 - v_2) \cdots (1 - v_{i-1})\}\left(1 - \frac{w}{B}\right) < 1$$

for $w > 0$ and $0 \leqq v_i \leqq 1$ $(i = 1, 2, \cdots, q - 1)$.

We denote the left hand side of (5) by ζ_i. The factor $(1 - \zeta_i)^{b_i - c_i + 1}$ can be expanded in a power series, i.e.

$$(6) \quad (1 - \zeta_i)^{b_i - c_i + 1} = (1 - \zeta_i)^{-(c_{i+1} - b_i)}$$
$$= 1 + (c_{i+1} - b_i)\zeta_i + \tfrac{1}{2}(c_{i+1} - b_i)(c_{i+1} - b_i + 1)\zeta_i^2 + \cdots$$

with a radius of convergence equal to one Since we will show shortly that for the choices we make for the b_i's and c_i's, $c_{i+1} \geq b_i$, then all coefficients in this last expansion are non-negative. Substituting this series expansion (6) in (4), and ordering it according to powers of $(1 - w/B)$, the expression under the integral sign (in 4) becomes

$$(7) \quad \theta_0(v_1, v_2, \cdots v_{q-1}) + \theta_1(v_1, \cdots v_{q-1})\left(1 - \frac{w}{B}\right) + \theta_2(v_1, \cdots, v_{q-1})\left(1 - \frac{w}{B}\right)^2 + \cdots.$$

This series is uniformly convergent over the domain defined by the inequalities $0 \leqq v_i \leqq 1$ $(i = 1, 2, \cdots, q - 1)$ and $| 1 - w/B | < 1$. We can even say that (7) is uniformly convergent for $| 1 - w/B | < 1$ if we substitute for each θ_i the maximum of θ_i with respect to $v_1, v_2, \cdots, v_{q-1}$. Hence we may integrate the series (7) with respect to $v_1, v_2, \cdots v_{q-1}$ term by term, i.e.

$$(8) \quad \int_0^1 \int_0^1 \cdots \int_0^1 (7)\, dv_1\, dv_2 \cdots dv_{q-1} = \sigma_0 + \sigma_1\left(1 - \frac{w}{B}\right) + \sigma_2\left(1 - \frac{w}{B}\right)^2 + \cdots$$

and the series (8) is uniformly convergent for $| 1 - w/B | < 1$. The coefficients $\sigma_0, \sigma_1, \cdots$ are non-negative.

The case of the λ statistic which we are considering is a special case of this integral equation which we obtain by making the following substitutions:

$$w = \lambda, \quad B = 1, \quad u = r + p_1, \quad q = p - p_1$$

$$b_r = \tfrac{1}{2}(n - u + 1), \quad c_r = \tfrac{1}{2}(n - u + u' + 1), \quad (r = 1, 2, \cdots, p - p_1)$$

Note that then

$$c_{r+1} - b_r = \tfrac{1}{2}[(u + 1)' - 1] \geqq 0.$$

Hence, according to (4)

$$g(\lambda)\, d\lambda = k \cdot \lambda^{\frac{1}{2}(n-p-1)}(1 - \lambda)^{\frac{1}{2}\Sigma u' - 1}\{\sigma_0 + \sigma_1(1 - \lambda) + \sigma_2(1 - \lambda^2) + \cdots\}\, d\lambda$$

where the infinite series converges for $| 1 - \lambda | < 1$.

Now $v = -\log \lambda$, or $\lambda = e^{-v}$, hence

$$h(v)\, dv = k \cdot e^{-\frac{1}{2}(n-p+1)v} v^{r-1}\left(\frac{1 - e^{-v}}{v}\right)^{r-1}\{\epsilon_0 + \epsilon_1 v + \epsilon_2 v^2 + \cdots\}\, dv$$

where the series $\{\epsilon_0 + \epsilon_1 v + \epsilon_2 v^2 + \cdots\}$ is obtained from the series $\{\sigma_0 + \sigma_1(1-\lambda) + \cdots\}$ by substituting for $(1-\lambda)$ the Taylor expansion of $(1 - e^{-v})$. The series $\{\epsilon_0 + \epsilon_1 v + \epsilon_2 v^2 + \cdots\}$ has a finite radius of convergence.[5]

Hence the function $\psi(v)$ can be written as

$$\psi(v) = A \cdot e^{\frac{1}{2}(p-1)v}\left(\frac{1-e^{-v}}{v}\right)^{r-1}\{\epsilon_0 + \epsilon_1 v + \epsilon_2 v^2 + \cdots\}$$

where A denotes a constant factor. Then since $e^{\frac{1}{2}(p-1)v}\left(\frac{1-e^{-v}}{v}\right)^{r-1}$ can be expanded in a Taylor series around $v = 0$, Proposition 4 is proved.

4. **Evaluation of the coefficients in the expansion of $\psi(v)$.** Let the series expansion of $\psi(v)$ be

$$\psi(v) = \alpha_0 + \alpha_1 v + \alpha_2 v^2 + \cdots$$

Then we have

$$\int_0^\infty \frac{c_n e^{-\frac{1}{2}nv} v^{r-1}}{\Gamma(r)}(\alpha_0 + \alpha_1 v + \alpha_2 v^2 + \cdots)\,dv \equiv 1.$$

Now let $v^* = \frac{n}{2}v$, then

$$\int_0^\infty \left(\frac{2}{n}\right)^r \frac{c_n e^{-v^*} v^{*r-1}}{\Gamma(r)}\left(\alpha_0 + \frac{2\alpha_1 v^*}{n} + \frac{4\alpha_2 v^{*2}}{n^2} + \cdots\right)dv^* \equiv 1.$$

Suppose that the asymptotic expansion of $\left(\frac{n}{2}\right)^r \frac{1}{c_n}$ is given by

$$\beta_0 + \frac{\beta_1}{n} + \frac{\beta_2}{n^2} + \cdots.$$

On account of Proposition 3, we have that the asymptotic expansion in powers of $1/n$ of

$$\int_0^{\sqrt{n}} \frac{e^{-v^*} v^{*r-1}}{\Gamma(r)}\left(\alpha_0 + \frac{2\alpha_1}{n}v^* + \frac{4\alpha_2}{n^2}v^{*2} + \cdots\right)dv^* \tag{9}$$

must be equal to the asymptotic expansion of $\left(\frac{n}{2}\right)^r \frac{1}{c_n}$. Since we may integrate in (9) term by term for sufficiently large n, we easily obtain

$$\alpha_0 = \beta_0, \qquad \alpha_1 = \frac{\beta_1}{2r}, \cdots \alpha_k = \frac{\beta_k}{2^k \cdot r(r+1)\cdots(r+k-1)}.$$

[5] See A. Gutzmer, *Theorie der Eindeutigen Analytischen Funktionen*, 1906, pp. 91–2.

The asymptotic expansion of $\left(\frac{n}{2}\right)^r \frac{1}{c_n}$ can be calculated in the following manner

$$\left(\frac{n+2}{n}\right)^r \frac{c_n}{c_{n+2}} = \frac{\beta_0 + \dfrac{\beta_1}{n+2} + \dfrac{\beta_2}{(n+2)^2} + \cdots}{\beta_0 + \dfrac{\beta_1}{n} + \dfrac{\beta_2}{n^2} + \cdots}$$

and

$$\left(\frac{n+2}{n}\right)^r \frac{c_n}{c_{n+2}} = (1 + 2/n)^r \prod_u \frac{n - u + 1}{n - u + u' + 1}$$

Equating the right hand members of these last two equations, and taking logs, we obtain

$$\log\left[\beta_0 + \frac{\beta_1}{n+2} + \frac{\beta_2}{(n+2)^2} + \cdots\right] = r \log (1 + 2/n) + \sum_u \log\left(1 - \frac{u-1}{n}\right) - \sum_u \log\left(1 - \frac{u - u' - 1}{n}\right) + \log\left(\beta_0 + \frac{\beta_1}{n} + \frac{\beta_2}{n^2} + \cdots\right).$$

Then we expand each term in a series of powers of $1/n$ and equate coefficients of $1/n^i$ for each i. We obtain the following formulae for the first five β's:

$$\beta_0 = 1$$

$$\beta_1 = r + \tfrac{1}{4} \sum_u (u-1)^2 - \tfrac{1}{4} \sum_u (u - u' - 1)^2$$

$$\beta_2 = \beta_1 + \frac{\beta_1^2}{2} - \frac{2r}{3} + \frac{1}{12} \sum_u (u-1)^3 - \frac{1}{12} \sum_u (u - u' - 1)^3$$

$$\beta_3 = -\tfrac{4}{3}\beta_1 - \beta_1^2 - \tfrac{1}{3}\beta_1^3 + \beta_1\beta_2 + 2\beta_2 + \tfrac{2}{3}r + \tfrac{1}{24} \sum_u (u-1)^4 - \tfrac{1}{24} \sum_u (u - u' - 1)^4$$

$$\beta_4 = 2\beta_1 + 2\beta_1^2 + \beta_1^3 + \frac{\beta_1^4}{4} - 3\beta_1\beta_2 + \beta_1\beta_3 - \beta_1^2\beta_2 - 4\beta_2 + \frac{\beta_2^2}{2} + 3\beta_3 - \frac{4}{5}r + \frac{1}{40} \sum_u (u-1)^5 - \frac{1}{40} \sum_u (u - u' - 1)^5.$$

5 **Practical use of the series.** In practical applications, the value of the statistic, say λ_0, is calculated, and it is desired that we determine whether or not this value of the statistic falls into the critical region. That is, for a particular grouping of the variates, for a particular number of degrees of freedom, and for a chosen level of significance α, there is determined from the distribution of λ, a value λ^* such that

$$\text{Prob}\,[\lambda < \lambda^*] = \alpha,$$

and if $\lambda_0 < \lambda^*$ we reject the hypothesis that in the population from which the sample is taken all the correlation coefficients between variates in different groups are zero.

Since v is a monotonic decreasing function of λ we make the test by computing $v_0 = -\log \lambda_0$ and we reject the hypothesis if $v_0 > v^*$ where $v^* = -\log \lambda^*$. But this is equivalent to computing Prob $[v > v_0]$ and if this value is less than α we reject the hypothesis Now

$$\text{Prob } [v > v_0] = J_{v_0}(r_1, r_2, \cdots, r_k)$$
$$= \frac{C_n}{\Gamma(r)} \int_{v_0}^{\infty} e^{-\frac{1}{2}nv} v^{r-1}(1 + \alpha_1 v + \alpha_2 v^2 + \cdots)\, dv$$

Setting $\frac{nv}{2} = z$

$$\text{Prob } [v > v_0] = \left(\frac{2}{n}\right)^r \frac{C_n}{\Gamma(r)} \int_{nv_0/2}^{\infty} e^{-z} z^{r-1} \left[1 + \alpha_1 \frac{2z}{n} + \alpha_2 \left(\frac{2}{n}\right)^2 z^2 + \cdots\right] dz.$$

On account of Proposition 3 we obtain an asymptotic expansion of Prob $[v > v_0]$ by integrating the right hand member of the above equation term by term. This can be expressed by means of the incomplete gamma function, which is tabulated[6] in the form

$$I(u, p) = \frac{\int_0^{u\sqrt{p+1}} v^p e^{-v}\, dv}{\Gamma(p+1)}.$$

We obtain

$$\text{Prob } [v > v_0] = \left(\frac{2}{n}\right)^r c_n \left\{\left[1 - I\left(\frac{nv_0}{2\sqrt{r}}, r - 1\right)\right]\right.$$
$$\left. + \frac{\beta_1}{n}\left[1 - I\left(\frac{nv_0}{2\sqrt{r+1}}, r\right)\right] + \frac{\beta_2}{n^2}\left[1 - I\left(\frac{nv_0}{2\sqrt{r+2}}, r+1\right)\right] + \cdots\right\}.$$

The values of the constant $K = \left(\frac{2}{n}\right)^r c_n$ and the values of β_1, β_2, β_3, β_4 are herein tabulated for any grouping which might be made on six or fewer variates. Some cases, such as groupings $(1, p-1)$, in which case the distribution of λ is the distribution of the multiple correlation coefficient; and as the groupings $(2, p-2)$, the exact distribution for which was given by Wilks as an incomplete Beta-function, are superfluous here. These cases are included only for the sake of completeness

[6] K. Pearson (Editor), *Tables of the Incomplete Gamma Function*, Biometric Laboratory, London, 1922.

Table of the First Four β's

Grouping	r	β_1	β_2	β_3	β_4
2,1	1	2	4	8	16
1,1,1	1.5	2.75	6.28125	13.38281	27.57568
3,1	1.5	3.75	12.03125	36 91406	111 55225
2,2	2	5	19	65	211
2,1,1	2.5	5.75	23.53125	83.97656	279.50538
1,1,1,1	3	6.5	28.625	106.9375	366.39844
4,1	2	6	28	120	496
3,2	3	9	55	285	1351
3,1,1	3.5	9.75	62.53125	334.10156	1615.91163
2,2,1	4	11	77	439	2229
2,1,1,1	4.5	11.75	86.03125	506.16406	2628.23974
1,1,1,1,1	5	12.5	95 625	580.6875	3085.52344
5,1	2.5	8.75	55.78125	315.82031	1690.65282
4,2	4	14	125	910	5901
3,3	4.5	15.75	154.03125	1205.03906	8277.55226
4,1,1	4.5	14.75	136.28125	1015.50781	6693.45068
3,2,1	5.5	17.75	189.53125	1584.10156	11445.75538
2,2,2	6	19	214	1866	13947
3,1,1,1	6	18.5	203.625	1740.9375	12797.27344
2,2,1,1	6.5	19.75	229.03125	2042.16406	15530.08351
2,1,1,1,1	7	20.5	244.625	2230.1875	17257.64836
1,1,1,1,1,1	7.5	21.25	260 78125	2430.49219	19139.02892

Tables of the Constant $K = \binom{2}{n}^r C_n$

n	21	111	31	22	211	1111	41	311
10	.800	.738	.646	.560	.517	.477	.480	.310
11	.818	.761	.676	.595	.553	.515	.521	.352
12	.833	.780	.702	.625	.585	.548	.556	.390
13	.846	.796	.724	.651	.612	.576	.586	.424
14	.857	.810	.743	.674	.637	.602	.612	.455
15	.867	.822	.759	.693	.658	.624	.636	.482
16	.875	.833	.774	.711	.677	.645	.656	.508
17	.882	.843	.787	.727	.694	.663	.675	.531
18	.889	.851	.798	.741	.709	.679	.691	.552
19	.895	.859	.808	.754	.723	.694	.706	.571
20	.900	.866	.818	.765	.736	.708	.720	.589
22	.909	.878	.834	.785	.758	.732	.744	.620
24	.917	.888	.847	.802	.777	.752	.764	.647
26	.923	.896	.859	.817	.793	.770	.781	.671
28	.929	.903	.869	.829	.807	.785	.796	.691
30	.933	.910	.877	.840	.819	.798	.809	.710
35	.943	.922	.894	.862	.843	.825	.835	.747
40	.950	.932	.908	.879	.862	.846	.855	.776
45	.956	.940	.918	.892	.877	.862	.871	.799
50	.960	.946	.926	.902	.889	.875	.883	.818
55	.964	.950	.932	.911	.899	.886	.894	.833
60	.967	.954	.938	.918	.907	.895	.902	.846
65	.969	.958	.943	.924	.914	.903	.910	.858
70	.971	.961	.947	.930	.920	.910	.916	.867
80	.975	.966	.953	.938	.930	.921	.926	.883
90	.978	.970	.959	.945	.937	.929	.934	.896
100	.980	.973	.963	.951	.943	.936	.941	.906

Tables of the Constant K (u)

n	221	2111	32	11111	51	42	33
10	.269	.248	.336	.229	.323	.168	.136
11	.310	.288	.379	.268	.369	.206	.171
12	.347	.325	.417	.304	.410	.243	.205
13	.381	.359	.451	.338	.445	.277	.237
14	.412	.390	.481	.368	.478	.309	.268
15	.441	418	.508	.397	.506	.339	.297
16	.467	.444	.533	.423	.532	.367	.324
17	.490	.468	.556	.447	.555	.392	.350
18	.512	.490	.576	.470	.576	.416	.374
19	.532	.511	.595	.490	.596	.438	.396
20	.551	.530	.612	.510	.613	.459	.417
22	.584	.564	.642	.544	.644	.496	.455
24	.613	.593	.668	.575	.671	.529	.489
26	.638	.619	.691	.601	.694	.558	.519
28	.660	.642	.711	.625	.714	.584	.546
30	.680	.662	.728	.646	.731	.607	.570
35	.720	.704	.764	.689	.767	.654	.621
40	.751	.737	.791	.723	.794	.692	.661
45	.776	.763	.813	.751	.816	.722	.694
50	.797	.785	.830	.773	.833	.747	.721
55	.814	.803	.845	.792	.848	.768	.743
60	.828	.818	.857	.808	.860	.786	.762
65	.841	.831	.868	.822	.870	.801	.779
70	.852	.842	.877	.833	.879	.814	.793
80	.869	.861	.892	.853	.894	.836	.817
90	.883	.876	.903	.869	.905	.853	.836
100	.894	.888	.913	.881	.915	.867	.852

Tables of the Constant K (iii)

n	411	321	222	3111	2211	21111	111111
10	.155	.108	.094	.100	.087	.080	.076
11	.192	.140	.123	.130	.114	.106	.099
12	.228	.171	.152	.160	.142	.133	.125
13	.261	.201	.180	.189	.170	.160	.150
14	.292	.230	.208	.217	.197	.186	.176
15	.322	.257	.235	.244	.223	.212	.201
16	.349	.284	.261	.270	.248	.236	.225
17	.375	.309	.285	.295	.272	.260	.248
18	.398	.332	.308	.318	.295	.283	.271
19	.421	.354	.330	.340	.317	.304	.292
20	.442	.375	.351	.361	.338	.325	.313
22	.479	.414	.390	.400	.376	.363	.351
24	.512	.448	.424	.434	.411	.398	.385
26	.542	.479	.456	.465	.442	.430	.417
28	.568	.507	.484	.493	.471	.458	.446
30	.591	.532	.510	.519	.497	.484	.472
35	.640	.585	.564	.573	.552	.540	.528
40	.679	.628	.608	.616	.597	.585	.574
45	.710	.663	.644	.652	.633	.623	.612
50	.736	.692	.674	.681	.664	.654	.644
55	.758	.716	.700	.706	.690	.681	.671
60	.776	.737	.722	.728	.712	.704	.695
65	.792	.755	.740	.746	.732	.723	.715
70	.805	.771	.757	.762	.749	.741	.733
80	.828	.797	.784	.789	.777	.770	.762
90	.846	.818	.806	.811	.800	.793	.786
100	.860	.835	.824	.828	.818	.812	.806

THE MEAN SQUARE SUCCESSIVE DIFFERENCE

By J. von Neumann,[1] R. H. Kent, H. R. Bellinson and B. I. Hart

Aberdeen Proving Ground

1. Introduction. In making measurements, every precaution is generally taken to hold the conditions of the experiment constant, in order that the population, whose parameters are to be estimated from the observations, shall remain fixed throughout the experiment One wishes each observation to come from the same population, or what is the same thing if normality is assumed, from populations having the same means and standard deviations.

There are cases, however, where the standard deviation may be held constant, but the mean varies from one observation to the next. If no correction is made for such variation of the mean, and the standard deviation is computed from the data in the conventional way, then the estimated standard deviation will tend to be larger than the true population value. When the variation in the mean is gradual, so that a trend (which need not be linear) is shifting the mean of the population, a rather simple method of minimizing the effect of the trend on dispersion is to estimate standard deviation from differences. It is for this purpose that the mean square successive difference

$$\delta^2 = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{n-1} \tag{1}$$

is suggested. The subscript i in this expression refers to the temporal order of the observation x_i.

In using δ^2 for estimating standard deviation, the distribution of δ^2 in random samples is of interest, since questions of bias, efficiency, and confidence interval require consideration. δ^2 may be used, in addition, to determine whether a trend actually exists; in this case one must know whether δ^2 differs significantly from

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}, \tag{2}$$

which measures variance independently of the order of the observations, and consequently includes the effect of the trend.

[1] Institute for Advanced Study, Princeton, N. J. Also member of Scientific Advisory Committee of the Ballistic Research Laboratory, Aberdeen Proving Ground.

The distribution of δ^2 is considered in this paper; it is hoped that others will shortly publish methods of estimating the probability that $\delta^2 \leqq ks^2$ as a function of k and the sample size n.

2. History. A somewhat similar procedure is suggested by "Student" [1] and E. S. Pearson [2] who consider the situation in which a shift may occur in the mean of the population, but where pairs of observations may be made with no shift in mean between them; standard deviation may be estimated from the differences between these pairs. The method can be generalized, and

$$s' = \sqrt{\frac{\sum_{i=1}^{n/2} (x_{2i} - x_{2i-1})^2}{n}}$$

is an estimate of the standard deviation. n must, of course, be an even integer. This estimate has the advantage that its properties are fully known: s' is distributed as the standard deviation with $f = n/2$ degrees of freedom. It will be noted that this estimate does not involve the successive differences, but only the alternate ones. Although there are $n - 1$ available successive differences, this estimate uses only the $n/2$ independent differences. The mean square successive difference is based on all $n - 1$ successive differences, and should therefore provide a more efficient estimate of σ than does s'.

There is, of course, nothing new in the concept of estimating the standard deviation from differences Even as far back as 1870, an interest in the method appears to have existed. Jordan [3] devised methods based on sums of powers of the differences. Helmert [4] gave more careful consideration to the case of the first power, i.e. the sum of the absolute differences. In both these cases, however, all the $n(n - 1)/2$ differences that can be established from a sample of n observations were included in the estimate, so that the estimate was of no value in reducing the effect of a trend. Helmert realized this, for he pointed out that the estimate obtained from the sum of squares of the differences is exactly that obtained by the more conventional procedure of squaring deviations from the mean.

The usefulness of the differences between successive observations only appears to have been realized first by ballisticians, who faced the problem of minimizing effects due to wind variation, heat and wear in measuring the dispersion of the distance traveled by shell. Vallier [5] appears to have been the first to estimate dispersion from successive differences. Cranz and Becker [6] commended the mean successive difference

$$E_d = \frac{\sum_{i=1}^{n-1} |x_{i+1} - x_i|}{n - 1}.$$

To establish the precision of E_d in estimating σ, Cranz and Becker quoted Helmert's paper, and so erred in saying that their method was superior to that

of the mean deviation. Helmert's procedure, based on $n(n-1)/2$ differences, is indeed more precise (for $n > 10$) than the mean deviation

$$M.D. = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n},$$

but the mean successive difference is based on but $n-1$ differences, and so is not as precise.

Bennett [7] appears to have suggested the use of successive differences independently of the European ballisticians. In recent years, the method of estimation by the mean square successive difference δ^2 was put into practice in the Ballistic Research Laboratory at the Aberdeen Proving Ground, U. S. Army, by L. S. Dederick

3. Bias and efficiency. The moments of δ^2 in samples drawn from a normal population are derived in Section 6 of this paper. The moments are used at this point to establish the estimate of variance, and the efficiency of this estimate.

The mean value of δ^2 in samples taken at random from a normal population is

$$E(\delta^2) = 2\sigma^2. \tag{3}$$

δ^2 consequently offers an unbiased estimate of variance, and this estimate is

$$\frac{\delta^2}{2} = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{2(n-1)}. \tag{4}$$

The second moment, i.e., the variance, of δ^2 in samples of size n:

$$\sigma_{\delta^2}^2 = \frac{4(3n-4)}{(n-1)^2}\sigma^4. \tag{5}$$

As the sample size is increased, the distribution of δ^2 appears to approach the normal. It is therefore appropriate to consider the efficiency as defined by Fisher [8]. Accordingly, the efficiency of δ^2 is

$$\left[\frac{\sigma_{s^2}}{E(s^2)} \Big/ \frac{\sigma_{\delta^2}}{E(\delta^2)}\right]^2.$$

Since

$$\sigma_{s^2}^2 = \frac{2(n-1)}{n^2}\sigma^4,$$

and

$$E(s^2) = \frac{n-1}{n}\sigma^2,$$

the efficiency of δ^2 in estimating the standard deviation is

$$\frac{2(n-1)}{3n-4} = \frac{2}{3}\left[1 + \frac{1}{3n-4}\right]. \tag{6}$$

The efficiency is unity for $n = 2$, since in this case the two statistics have the same distribution. It therefore appears that the efficiency decreases as the sample size increases, but approaches 2/3 as a limiting value for n very large.

4. Summary of procedure. Having a statistic which estimates a parameter of a population, it is desirable to know the distribution of that statistic as computed from samples taken at random from that population At present, the distribution of δ^2 in samples of n has not been obtained. The difficulty is in the fact that the successive differences are not independent. The first difference, $d_1 = x_2 - x_1$, and the second difference, $d_2 = x_3 - x_2$, are related in that they both involve x_2. Similar correlation exists between every successive pair of differences between successive observations.

For $n = 2$, and samples taken from a normal population, the distribution of δ^2 is known. Since

$$\delta^2 = (x_2 - x_1)^2 = 2\sum_{i=1}^{2}(x_i - \bar{x})^2 = 4s^2,$$

the distribution of δ^2 is similar to that of s^2 for this sample size.

For $n = 3$, the distribution of δ^2 has been derived analytically. The derivation is indicated in Section 5 of this paper. For $n > 3$, only the moments of the distribution have thus far been obtained. A Pearson type distribution has been fitted to the first three moments to obtain an approximate representation of the true distribution.

5. Distribution of δ^2. In the case of a sample of n taken from a normal population, the probability that the first observation lies between x_1 and $x_1 + dx_1$, while the second lies between x_2 and $x_2 + dx_2$, etc., is

$$\left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n e^{-(x_1^2+x_2^2+\cdots+x_n^2)/2\sigma^2}\, dx_1 dx_2 \cdots dx_n. \tag{7}$$

If $y_i = x_{i+1} - x_i$, this expression becomes

$$\left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n e^{-Q(x_1, y_1, y_2, \cdots, y_{n-1})/2\sigma^2}\, dx_1\, dy_1\, dy_2 \cdots dy_{n-1}, \tag{8}$$

where Q is a quadratic form in x_1 and the y's. Since

$$\delta^2 = \frac{\sum_{i=1}^{n-1} y_i^2}{n-1},$$

the probability that δ^2 shall be less than some value δ_0^2 is

$$(9)\quad P(\delta^2 < \delta_0^2) = \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n \underset{\sum_{i=1}^{n-1} y_i^2 < (n-1)\delta_0^2}{\int\int \cdots \int} \int_{-\infty}^{+\infty} e^{-Q(x_1, y_1, \cdots y_{n-1})/2\sigma^2}\, dx_1\, dy_1 \cdots dy_{n-1}.$$

After the integration with respect to x_1 is carried out, the quadratic form in the exponent may be normalized by a transformation to new coordinates z_i linearly related to the y's. The z's may be so chosen that all the terms z_i^2 in the exponent have the same coefficient, in which case

$$(10)\quad P(\delta^2 < \delta_0^2) = c_1 \int\int \cdots \int e^{-k\sum_{i=1}^{n-1} z_i^2} \frac{\partial(y_1, y_2, \cdots, y_{n-1})}{\partial(z_1, z_2, \cdots, z_{n-1})}\, dz_1\, dz_2 \cdots dz_{n-1}.$$

As a result of such a transformation, the sphere of integration in (9) becomes an ellipsoid in (10). By changing to polar coordinates, with

$$(11)\quad \begin{aligned} r^2 &= \sum_{i=1}^{n-1} z_i^2, \\ P(\delta^2 < \delta_0^2) &= c_1 \int\int e^{-kr^2} r^{n-2}\, d\Omega\, dr, \end{aligned}$$

in which Ω is the solid angle in the space of $n - 1$ dimensions. The limits of integration with respect to Ω as a function of r must be found; this involves the evaluation of the solid angle subtended by the surface bounded by the intersection of the $(n - 1)$-dimensional sphere and the $(n - 1)$-dimensional ellipsoid. If $\Omega = \phi(r)$,

$$(12)\quad P(\delta^2 < \delta_0^2) = c_2 \int_0^a e^{-kr^2} \phi(r) r^{n-2}\, dr,$$

in which a is the longest semi-axis of the $(n - 1)$-dimensional ellipsoid corresponding to the given value of δ^2.

For $n = 3$, (9) becomes

$$(13)\quad \begin{aligned} P(\delta^2 < \delta_0^2) &= \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^3 \underset{y_1^2 + y_2^2 < 2\delta_0^2}{\int\int} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{3\sigma^2}(y_1^2 + y_2^2 + y_1 y_2)\right. \\ &\qquad \left. - \frac{3}{2\sigma^2}\left(x_1 + \frac{2y_1 + y_2}{3}\right)^2\right] dx_1\, dy_1\, dy_2 \\ &= \frac{1}{2\sqrt{3}\,\pi\sigma^2} \underset{y_1^2 + y_2^2 < 2\delta_0^2}{\int\int} e^{-(y_1^2 + y_1 y_2 + y_2^2)/3\sigma^2}\, dy_1\, dy_2. \end{aligned}$$

Normalizing the quadratic form in the exponent,

$$(14)\quad P(\delta^2 < \delta_0^2) = \frac{1}{2\sqrt{3}\,\pi\sigma^2} \underset{z_1^2 + z_2^2 < 2\delta_0^2}{\int\int} e^{-(z_1^2 + \frac{1}{4}z_2^2)/2\sigma^2}\, dz_1\, dz_2,$$

and in polar coordinates

$$P(\delta^2 < \delta_0^2) = \frac{1}{2\sqrt{3}\,\pi\sigma^2}\int_0^{\delta_0\sqrt{2}}\int_0^{2\pi} re^{-r^2[\cos^2\theta + \frac{1}{3}\sin^2\theta]/2\sigma^2}\,d\theta\,dr$$

$$= \frac{1}{2\sqrt{3}\,\pi\sigma^2}\int_0^{\delta_0\sqrt{2}} re^{-r^2/2\sigma^2}\left[\int_0^{2\pi} e^{r^2\sin^2\theta/3\sigma^2}\,d\theta\right]dr. \tag{15}$$

The integral in brackets can be shown to be a Bessel function of zero order; for let

$$r^2/3\sigma^2 = -2iu,$$

$$\phi = \frac{\pi}{2} - 2\theta,$$

then

$$\int_0^{2\pi} e^{r^2\sin^2\theta/3\sigma^2}\,d\theta = e^{-iu}\int_{-\pi}^{\pi} e^{iu\sin\phi}\,d\phi = 2\pi e^{-iu}J_0(u). \tag{16}$$

Consequently, (15) takes the form

$$P(\delta^2 < \delta_0^2) = \frac{1}{\sigma^2\sqrt{3}}\int_0^{\delta_0\sqrt{2}} re^{-r^2/3\sigma^2}J_0\left(\frac{ir^2}{6\sigma^2}\right)dr = F(\delta_0^2). \tag{17}$$

The probability density function

$$p(\delta^2) = \frac{dF(\delta^2)}{d\delta^2}$$

$$= \frac{1}{\sigma^2\sqrt{3}}e^{-2\delta^2/3\sigma^2}J_0\left(\frac{i\delta^2}{3\sigma^2}\right) \tag{18}$$

$$= \frac{1}{\sigma^2\sqrt{3}}e^{-2\delta^2/3\sigma^2}\left[1 + \frac{1}{2^2}\frac{\delta^4}{3^2\sigma^4} + \frac{1}{2^2 4^2}\frac{\delta^8}{3^4\sigma^8} + \frac{1}{2^2 4^2 6^2}\frac{\delta^{12}}{3^6\sigma^{12}} + \cdots\right].$$

6. Moments. The t-th moment of δ^2 about the origin is defined by

$$\mu_t' = E[(\delta^2)^t], \tag{19}$$

or

$$(n-1)^t\mu_t' = E\left(\left[\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2\right]^t\right)$$

$$= E\left(\left[2\sum_{i=1}^{n}x_i^2 - (x_1^2 + x_n^2) - 2\sum_{i=1}^{n-1}x_{i+1}x_i\right]^t\right). \tag{20}$$

For any value of t, the expansion can be performed, and similar terms collected and enumerated. The values of x can be considered as true errors, i.e. as deviations from the true mean, without affecting the conclusions. If the

original population from which the samples have been drawn is normal, with standard deviation σ, then

$$E(x^{2k-1}) = 0$$
$$E(x^{2k}) = \frac{(2k)!}{2^k k!}\sigma^{2k}, \tag{21}$$

and since, in the null case where the mean of the population remains constant, successive observations are independent, then

$$E(x_i^r x_j^s) = E(x^{r+s}), \qquad i = j$$
$$E(x_i^r x_j^s) = E(x^r)E(x^s), \qquad i \neq j. \tag{22}$$

These relations are sufficient for the evaluation of μ_t'. For example, in the case of the second moment, $t = 2$:

$$(n-1)^2\mu_2' = E\left(\left[2\sum_{i=1}^{n} x_i^2 - (x_1^2 + x_n^2) - 2\sum_{i=1}^{n-1} x_{i+1}x_i\right]^2\right). \tag{23}$$

Now:

$$\left[2\sum_{i=1}^{n} x_i^2 - (x_1^2 + x_n^2) - 2\sum_{i=1}^{n-1} x_{i+1}x_i\right]^2$$
$$= 4\left(\sum_{i=1}^{n} x_i^2\right)^2 + (x_1^2 + x_n^2)^2 + 4\left(\sum_{i=1}^{n-1} x_{i+1}x_i\right)^2$$
$$- 4(x_1^2 + x_n^2)\sum_{i=1}^{n} x_i^2 - 8\sum_{i=1}^{n} x_i^2\sum_{i=1}^{n-1} x_{i+1}x_i + 4(x_1^2 + x_n^2)\sum_{i=1}^{n-1} x_{i+1}x_i$$
$$= 4\left[\sum_{i=1}^{n} x_i^4 + \sum_{i,j=1, i\neq j}^{n} x_i^2 x_j^2\right] + [x_1^4 + 2x_1^2x_n^2 + x_n^4]$$
$$+ 4\left[\sum_{i=1}^{n-1} x_{i+1}^2 x_i^2\right] - 4\left[x_1^4 + x_1^2\sum_{i=2}^{n} x_i^2 + x_n^2\sum_{i=1}^{n-1} x_i^2 + x_n^4\right]$$
$$+ [\text{terms containing odd powers of } x_i].$$

The mean of these terms is found by using (21) and (22), and the number of each type of term present is enumerated:

$$4[n(3\sigma^4) + n(n-1)\sigma^2\sigma^2] + [3\sigma^4 + 2\sigma^2\sigma^2 + 3\sigma^4] + 4[(n-1)\sigma^2\sigma^2]$$
$$- 4[3\sigma^4 + \sigma^2(n-1)\sigma^2 + \sigma^2(n-1)\sigma^2 + 3\sigma^4] = (4n^2 + 4n - 12)\sigma^4.$$

Consequently

$$\mu_2' = \frac{4(n^2 + n - 3)}{(n-1)^2}\sigma^4. \tag{24}$$

The first four moments about the origin were evaluated by this procedure,

and from these, the moments about the mean are readily determined. The results are

$$
\begin{aligned}
\mu_1' &= 2\sigma^2 \\
\mu_2' &= \frac{4(n^2+n-3)}{(n-1)^2}\sigma^4 \\
\mu_3' &= \frac{8(n^3+6n^2+2n-21)}{(n-1)^3}\sigma^6 \\
\mu_4' &= \frac{16(n^4+14n^3+53n^2-8n-231)}{(n-1)^4}\sigma^8 \\
\mu_1 &= 0 \\
\mu_2 &= \frac{4(3n-4)}{(n-1)^2}\sigma^4 \\
\mu_3 &= \frac{32(5n-8)}{(n-1)^3}\sigma^6 \\
\mu_4 &= \frac{48(9n^2+46n-112)}{(n-1)^4}\sigma^8 .
\end{aligned}
\tag{25}
$$

It should be noted at this point that the above fourth moment is incorrect for $n = 2$. One of the terms in the expansion of the right side of (20), for $t = 4$, is

$$x_1^2 x_n^2 \sum_{i=1}^{n-1} x_{i+1}^2 x_i^2 .$$

For $n = 2$, the mean value of this term is

$$E(x_1^2 x_2^2 x_2^2 x_1^2) = E(x_1^4)E(x_2^4) = 9\sigma^8,$$

whereas for $n > 2$, the mean value is

$$E(x_1^4 x_2^2 x_n^2) + E\left(x_1^2 x_n^2 \sum_{i=2}^{n-2} x_{i+1}^2 x_i^2\right) + E(x_1^2 x_{n-1}^2 x_n^4) = (n+3)\sigma^8.$$

7. Pearson type fit to distribution of δ^2. From the moments it is found that

$$
\begin{aligned}
\beta_1 &= \frac{\mu_3^2}{\mu_2^3} = \frac{16(5n-8)^2}{(3n-4)^3}, \\
\beta_2 &= \frac{\mu_4}{\mu_2^2} = \frac{3(9n^2+46n-112)}{(3n-4)^2}.
\end{aligned}
\tag{26}
$$

As n becomes large, β_1 and β_2 approach 0 and 3 respectively; the distribution therefore appears to approach the normal for large samples. For finite sample sizes, the values of β_1 and β_2 correspond to those of the Pearson Type VI

distribution,

$$p\left(\frac{\delta^2}{\sigma^2}\right) = c\left(\frac{\delta^2}{\sigma^2} + a_1\right)^{q_2}\left(\frac{\delta^2}{\sigma^2} + a_2\right)^{-q_1}.$$

The origin of this distribution is at $\delta^2 = -a_1\sigma^2$, but the origin of the true distribution must be at $\delta^2 = 0$. By taking $a_1 = 0$ so that the origin is at $\delta^2 = 0$, we obtain what appears to be a suitable approximation

$$p\left(\frac{\delta^2}{\sigma^2}\right) = c\left(\frac{\delta^2}{\sigma^2}\right)^{q_2}\left(\frac{\delta^2}{\sigma^2} + a_2\right)^{-q_1}. \tag{27}$$

The parameters are determined by equating the 1st, 2nd and 3rd moments of (27) to the corresponding moments of the true distribution, with the result that

$$\begin{aligned} q_2 &= \frac{3n^4 - 10n^3 - 18n^2 + 79n - 60}{8n^3 - 50n + 48}, \\ q_1 &= \frac{4 - \mu_2(q_2 + 1)(q_2 + 3)}{4 - \mu_2(q_2 + 1)}, \\ a_2 &= \frac{2(q_1 - q_2 - 2)}{q_2 + 1}, \\ c &= \frac{a_2^{q_1 - q_2 - 1}}{\mathrm{B}(q_2 + 1, q_1 - q_2 - 1)}. \end{aligned} \tag{28}$$

Values of these parameters for selected values of n are given in Table I. The sixth and seventh columns of this table give the values of β_2 for the distribution (27) and for the true distribution, respectively.

TABLE I

(1) n	(2) q_1	(3) q_2	(4) a_2	(5) c	(6) β_2 (27)	(7) β_2 True	(8) Ratio (6)/(7)
5	24.4391	0.6391	26.6000	5.8800×10^{34}	8.807	8.504	1.036
7	31.1286	1.3857	23.2571	4.9285×10^{42}	6.948	6.758	1.028
10	41.2830	2.5079	20.9667	9.4934×10^{54}	5.658	5.538	1 022
15	58.2113	4.3806	19.2659	4.0240×10^{75}	4.718	4.645	1.016
20	75.1210	6.2543	18.4351	1.8063×10^{96}	4.269	4.217	1.012
25	92.0189	8.1285	17.9417	8.1097×10^{116}	4.006	3.965	1.010
50	176 4443	17.5018	16.9651	1.3386×10^{220}	3.494	3.475	1.005

The *Tables of the Incomplete Beta-Function* [9] can be used to evaluate the probability integral of the distribution (27),

$$\begin{aligned} P\left(\frac{\delta^2}{\sigma^2} < \frac{\delta_0^2}{\sigma^2}\right) &= c\int_0^{\delta_0^2/\sigma^2}\left(\frac{\delta^2}{\sigma^2}\right)^{q_2}\left(\frac{\delta^2}{\sigma^2} + a_2\right)^{-q_1} d\left(\frac{\delta^2}{\sigma^2}\right) \\ &= 1 - I_x(q_1 - q_2 - 1, q_2 + 1) \end{aligned} \tag{29}$$

$$x = \frac{a_2}{a_2 + \delta_0^2/\sigma^2},$$

for $n \leqq 14$. For $n > 14$, the probability integral may be determined by quadrature Some values of the probability integral for $n = 50$ are given in Table II. A comparison with the integral of the normal curve having the same first two moments indicates that a sample of somewhat more than 50 is required before the normal curve becomes a satisfactory approximation to the distribution (27).

TABLE II

$$P\left(\frac{\delta^2}{\sigma^2} < \frac{\delta_0^2}{\sigma^2}\right) \quad \text{for } n = 50$$

δ_0^2/σ^2	(29)	Normal
.50	·00000	·00118
.75	.00031	·00563
1.00	.00647	.02129
1.25	.04393	.06418

REFERENCES

[1] "STUDENT," *Biometrika*, Vol. 19(1927), p. 158.
[2] E. S. PEARSON, *Application of Statistical Methods to Industrial Standardisation and Quality Control*, London, 1935, p. 62
[3] W. JORDAN, *Astronomische Nachrichten*, Vol. 74(1869), pp. 209-226.
[4] F. R. HELMERT, *Astronomische Nachrichten*, Vol 88(1876), pp. 127–132.
[5] E. VALLIER, *Balistique Experimentale*, Paris, 1894, p. 166.
[6] C. CRANZ and K. BECKER, *Exterior Ballistics*, (trans. from 2nd German edition) London, 1921, p. 383
[7] A A BENNETT, unpublished report to the Chief of Ordnance, U. S. Army, circa 1918.
[8] R A FISHER, *Phil. Trans.* A, Vol 222(1922), p. 316
[9] KARL PEARSON (Editor), *Tables of the Incomplete Beta-Function*, London: Biometrika Office, 1934.

THE RETURN PERIOD OF FLOOD FLOWS

By E. J. Gumbel

New School for Social Research

Introduction. Engineers have used various interpolation formulas to represent the observed distribution of flood discharges. These formulas are sometimes constructed *ad hoc* for a given stream, and have no general meaning. Most of them are rather complicated.[1] Some authors have tried to introduce upper and lower limits to the discharges, even though it is doubtful that such limits exist. Others have introduced the third and fourth moments of the distribution, in spite of the fact that these numerical values are subject to large errors. For some formulas it is impossible to give a meaning to the constants; different formulas applied to the same stream give rather contradictory results; and consequently there is considerable confusion For example, Slade [20] has stated that "the statistical method in whatever form employed is an entirely inadequate tool in the determination of flood frequencies." According to Saville [19] "the engineer should satisfy himself that he has used an adequate number of methods, whether mathematical, graphic or otherwise, which have real support from either theory or experience, and then form his own judgement."

The main reason for this situation is that these studies have little or no theoretical basis. The author believes it possible to give exact solutions, exactitude being interpreted from the standpoint of the calculus of probabilities [10]. Our solutions are simply the consequences of a truism: "The flood discharges are the largest values of the discharges" The present study is but an explanation of this statement.

Many American authors start with a statistical function, which we call the return period of floods. Therefore we shall first analyse the notion of return period and show how it can be derived as a consequence of the concept of distribution. We then give a short résumé of the theory of largest values The discharge, and in consequence the flood discharge, is considered as an unlimited statistical variable; it is not necessary to determine its distribution. We are justified in representing the observed distribution of flows by one of the theoretical distributions of largest values. The distribution we choose contains only two constants, and both have a clear hydrological meaning. The numerical values are calculated by the method of moments.

---

[1] In recent years many articles discussing this topic have been published by the American Society of Civil Engineers and the American Geophysical Union [8]. A review of some of the proposed formulas is given in the Water Supply Paper 771 [17].

The application of the notion of return period to the largest values leads to a simple formula for the return period of the floods. In the last part of this paper we represent the flood flows of the Rhône and Mississippi Rivers by our formula.

1. The return period. Let us consider a continuous statistical variable x, having a theoretical distribution $w(x)$. The probability $W(x)$ of a value less than or equal to x, and the probability $P(x)$ of a value greater than or equal to x, are

$$W(x) = \int_{-\infty}^{x} w(z)\,dz, \qquad P(x) = \int_{x}^{\infty} w(z)\,dz, \tag{1}$$

where z denotes the variable of integration. Clearly

$$W(x) + P(x) = 1. \tag{1'}$$

Let n be the number of observations. Let x_m ($m = 1, 2, \cdots, n$) be the observed values arranged in increasing magnitude, where m is the serial number beginning with the lowest ("from below"). The lowest observation has the serial number $m = 1$, the highest has the serial number $m = n$. These observed values will be written x_1, and x_n respectively. The number of observations below or equal to x_m is $m = n'W(x_m)$ where $'W(x_m)$ is the observed relative number corresponding to the probability $W(x)$. The graphic representation of this series is called a cumulative histogram.

In hydraulics many authors arrange the observations in decreasing magnitude Let ${}_mx$ ($m = 1, 2, \cdots, n$) be these observed values. The serial number m is counted in a descending scale ("from above"). For the largest value $m = 1$, for the lowest value $m = n$. The number of observations above or equal to ${}_mx$ is $m = n'P({}_mx)$ where $'P({}_mx)$ corresponds to $P(x)$. The numbers $'W(x_m)$ will never decrease; the number $'P({}_mx)$ will never increase The mth value on a descending scale is the $n - m + 1$th value on an ascending scale. Therefore

$$n'P({}_mx) = n - n'W(x_m) + 1, \tag{2}$$

and

$$nP(x) = n - nW(x). \tag{2'}$$

The difference between formulas (2) and (2') will play a certain rôle later.

Different methods are used in statistics in comparing the theoretical values $W(x)$ or $P(x)$ and $w(x)$ with the corresponding observations $'W(x_m)$, or $'P({}_mx)$ (cumulative frequencies) and $\Delta'W(x_m)$ (frequency distribution). They all have in common an arrangement of observed values according to magnitude.

For the purpose of considering the observations in chronological order, we introduce a statistical criterion which at first glance may appear to have a new logical structure It is assumed here that the observations are made at constant time intervals, and this interval is considered the unit of time. We suppose that the observations are homogeneous, i e., subject to a common set of forces.

Furthermore, we suppose that the events are independent of one another. the occurrence of a high or low value for x has no influence on the value of any succeeding observation. Let us choose a low value x, and ask the following: After what number of observations does this or a greater value return? We calculate the mean of these chronological intervals between every two consecutive values, equal to or greater than x We repeat these operations for a second, third, . . . till the penultimate value of x.

These means are called the *observed return periods.* The criterion consists of the comparison of the observed, and the theoretical return period for increasing values of x. For a discontinuous variable we could obtain the return period for a value equal to x, (not equal to or greater than x). This average time, which is sometimes used in physics, does not interest us, as our variable, the discharge, is continuous. We limit our consideration to the return period of a value equal to or greater than x, called: value greater than x.

The determination of the theoretical return period is a classical problem: How many trials must, on the average, be made, in order that an event of a given probability should happen? Our event, the realization of a value, equal to or greater than x, has the probability $P(x) = 1 - W(x)$.

The mean number of trials $T(x)$ which are necessary to obtain our event once, is evidently

$$T(x) = \frac{1}{1 - W(x)}, \tag{3}$$

or

$$T(x) = \frac{1}{P(x)}. \tag{3'}$$

This value $T(x)$ is the mean chronological interval between two values, equal to or greater than x. If we start at the time when such a value has been observed for the first time, we can interpret $T(x)$ as the theoretical return period of a value equal to or greater than x. We designate it as the *theoretical return period.* This concept has not been used in statistics. It is a well-known concept in hydraulics which was introduced by Fuller [6]. To every theoretical distribution $w(x)$ there is a corresponding return period $T(x)$ and conversely, to every theoretical return period $T(x)$ there is a corresponding distribution

$$w(x) = \frac{T'(x)}{T^2(x)}, \tag{4}$$

obtained by differentiating (3).

If the variable is without limit to the left, the return period will start with $T = 1$. If the variable is limited to the left by $x \geqq \epsilon$ the corresponding return period will be

$$T(\epsilon) \geqq 1 \qquad \text{if } W(\epsilon) \geqq 0 \tag{5}$$

In the graphic representation, the return period $T(x)$ which has a time dimension, will be the abcissa and x the ordinate. Therefore we consider x as a function of $T(x)$; from (4) we obtain

$$\frac{dx}{d \ln T} = \frac{1}{w(x)T(x)} \tag{6}$$

where ln signifies the natural logarithm. The increase of x as a function of ln $T(x)$ will be very rapid for small values of T. For a limited distribution the same result is obtained, provided the probability $W(\epsilon)$ and the density of probability $w(\epsilon)$ are sufficiently small. Clearly, the return periods of the three quartiles are respectively $1\frac{1}{3}$, 2, 4. The return period will always increase with x. It will tend towards infinity even if the variable is limited to the right.

Let us now consider the calculus of the observed return periods. Instead of values equal to or greater than x_m we will only speak of values greater than x_m. The observed return period is the interval between the first and the last observation greater than x_m, divided by the number of intervals between all observations greater than x_m. The number of observations greater than x_m is $n - n'W(x_m)$. Between these observations there are $n - n'W(x_m) - 1$ intervals. This denominator is independent of the chronological order of the observed values. We can calculate the mean of the observed intervals up to a value x_m so that $n - n'W(x_m) = 2$ For this value of x_m there are only two observations, i.e., only one interval. In that case no mean can be calculated.

The numerator, the interval between the first and the last observation greater than x_m will be $n - 1$, provided that the first and the last value in chronological order are greater than x_m. But in general the first value greater than x_m will be the $('k + 1)$th in chronological order. The first value greater than x_m found in the reverse chronological order, will be the $(k' + 1)$th. Let $'k + k' = l$, then the interval between the last and the first value greater than x_m is $n - 1 - l$. The mean observed interval is thus

$$_1T(x_m) = (n - 1 - l)/(n - 1 - n'W(x_m)),$$

or

$$_1T(x_m) = \left(1 - \frac{l}{n-1}\right) \Big/ \left(1 - \frac{m}{n-1}\right). \tag{7}$$

This magnitude depends only on the chronological order of the first and the last value greater than x_m. It is independent of the chronological order of all other observations. Even in the case $l = 0$ this value differs from the theoretical value (3). The observed value surpasses the theoretical value, even if the frequency $'W(x_m)$ is identical with the probability $W(x)$.

In the general case, $l > 0$, this difference is a function of l. The number l depends upon the times at which the observations begin and cease; but it is not a characteristic of the chronological order. As a result of these disadvantages of formula (7) we prefer to introduce other definitions, in which the

chronological order does not enter. These definitions have an added advantage in that they are constructed in a manner analogous to the theoretical formula.

The observed value which corresponds to (3) is

$$'T(x_m) = \frac{n}{n - n'W(x_m)}, \tag{8}$$

or

$$'T(x_m) = n/(n - m). \tag{9}$$

But this definition of the observed return period is not the only one which corresponds to (3). Starting with the serial number m, in a descending scale, Fuller [6] puts

$$''T(x_m) = \frac{n}{m}. \tag{8'}$$

According to this definition, the return period of the mth value from below is

$$''T(x_m) = n/(n - m + 1). \tag{9'}$$

TABLE 1

Two definitions of the observed return periods

observed variable	serial number from below	serial number from above	exceedance interval formula (9)	recurrence interval formula (9′)
x_1	1	n	$n/(n-1)$	1
x_2	2	$n-1$	$n/(n-2)$	$n/(n-1)$
x_m	m	$n-m+1$	$n/(n-m)$	$n/(n-m+1)$
x_{n-1}	$n-1$	2	$n/1$	$n/2$
x_n	n	1	—	$n/1$

This observed return period corresponds to the theoretical return period (3′). The difference between (9) and (9′) results from the fact that the relation (2) between the observed cumulative frequencies $'W(x_m)$ and $'P(_mx)$ differs from the relation (2′) between the probabilities $W(x)$ and $P(x)$ The two definitions of the observed return periods are related by

$$''T(x_{m+1}) = 'T(x_m) < 'T(x_{m+1}). \tag{10}$$

From a purely logical standpoint the first definition is as justifiable as the second one.. Both are used in hydraulics. In order to avoid confusion between formulas (9) and (9′) Horton [16] calls $'T(x_m)$ the *exceedance* interval, i.e., "the average interval at which an event of given magnitude is exceeded," whereas he defines $''T(x_m)$, the recurrence interval as "the average interval of occurrence of values equalling or exceeding a given magnitude." Of course, the exceedance interval surpasses the recurrence interval. Since both observed intervals correspond to a common theoretical return period we designate both of them as observed return periods.

The difference between formulas (9) and (9′) is made clear in Table I

Each of the definitions (9) and (9′) and the theoretical expression $T(x)$ has different properties. For the lowest observation

$$n'W(x_1) = 1; \qquad n'P({}_nx) = n.$$

Therefore

$$'T(x_1) = 1 + \frac{1}{n-1}; \qquad ''T(x_1) = 1,$$

whereas for an unlimited distribution $\lim_{x \to -\infty} T(x) = 1$.

If the number of observations is sufficiently large the numerical differences between the two observed periods are rather small, except for very large values of the variable. For the last observation

$$n'W(x_n) = n; \qquad n'P({}_1x) = 1.$$

Therefore the return period $'T(x_n)$ for the last observation does not exist. According to the second definition the return period for the last value is equal to the total number of observations. But in general there is only one observation of the last value.

The preference given formula (9) over (9′) corresponds with the preference given to $W(x)$ over $P(x)$ when comparing the theoretical with the observed values. Therefore it is natural to count m from below. Since both definitions are equally applicable and since they lead to different results for large values of the variable, one should not calculate the return period for a small number of observations.

The observed return periods (9) and (9′) differ from the theoretical return period (3) in the same way that the frequencies $'W(x_m)$ or $'P({}_mx)$ differ from the probabilities $W(x)$ or $P(x)$. The chronological order enters neither into formula (7) nor into (9) or (9′). We need not take it into consideration, since the theoretical return period is obtained from the probability and the observed return period from the cumulative histogram. Therefore the usual statistical methods can be used for making the comparison between observed and theoretical return periods.

The return period is a statistical function like the distribution $w(x)$ or the probability $W(x)$. No formula for $T(x)$ that contradicts the properties of $w(x)$ can be accepted. The return period $T(x)$ will contain the same number of independent constants as the distribution $w(x)$. Consequently the fit of the theoretical curve $T(x)$ to the observations $'T(x_m)$ or $''T(x_m)$ cannot be improved by introducing a new constant without also changing the distribution $w(x)$. The theoretical curve $x = f(T)$ will fit the observed curves $(x_m, 'T(x_m))$ and $(x_m, ''T(x_m))$ in a way that depends upon the fit of $W(x)$ and $P(x)$ to $'W(x_m)$ and $'P({}_mx)$.

Let us suppose that $w(x)$ contains k constants; that they are determined by the method of moments which conserves the arithmetic mean $\bar{x}$, the mean of the squares $\overline{x^2}$ etc. of the observed distribution. For the return period these mo-

ments have a meaning. Let us consider for the sake of simplicity a positive variable. The kth moment M_k

$$M_k = \int_0^\infty x^k \, dW(x)$$

$$= -\int_0^\infty x^k \, d(1 - W(x))$$

$$= k \int_0^\infty (1 - W(x)) x^{k-1} \, dx$$

is according to (3)

$$M_k = k \int_0^\infty \frac{x^{k-1} \, dx}{T(x)}, \tag{11}$$

whence for $k = 1$ and $k = 2$

$$\bar{x} = \int_0^\infty \frac{dx}{T(x)}; \qquad E(x^2) = 2 \int_0^\infty \frac{x \, dx}{T(x)} \tag{11'}$$

For a given distribution containing two constants, the method of moments conserves the area and the center of gravity of the reciprocal of the return period. Even if the method of methods gives the best determination of the constants, for the distribution, it need not give the best determination for the return period. But if the observed return periods were used for the determination of the constants we would get two sets, since there are two observed curves having equal validity, but different values for large x. We will get one and only one set if the constants are calculated from the observed distribution, for here the difference between $'T(x_m)$ and $''T(x_m)$ does not matter The fact that we do not take the constants from the observed return periods, but from another statistical function, might be a cause for deviations between the observed and the theoretical return periods.

Once the constants have been found, we compare the observed curves $(x_m, 'T(x_m))$ and $(x_m, ''T(x_m))$ with the theoretical curve $x = f(T)$ To avoid discontinuity the observed return period will be established for all values of x_m arranged in increasing order.

If the observed return periods for small values of x are systematically smaller (greater) than the theoretical period, it is reasonable to conclude that there exists an attraction (repulsion) for small values of the variable and a repulsion (attraction) for the large values. But it must be remembered that the observed values have different weights in that the return periods for small values of x are based on many observations. This number diminishes as x increases. The last observed return period is based only on two observations Therefore the divergence between theory and observation will increase with the variable. With this precaution the criterion of the return period suggests one cause of difference between theory and observation. In order to apply this method to the largest values we must first establish the corresponding distribution.

2. Theory of the largest value. Let x be a statistical variable unlimited to the right having the distribution $w(x)$ Among the N observed values, one will be larger than the others. We wish to determine its theoretical value.

According to the principle of multiplication the probability $\mathfrak{W}_N(x)$ that N values are inferior to x is

$$\mathfrak{W}_N(x) = W^N(x). \tag{12}$$

This is the probability of x being the largest value. The largest value is a new statistical variable which possesses a mode, a mean $\bar{u}$, a standard deviation s and higher moments. To get the mean the distribution $\mathfrak{w}_N(x)$ of the largest value is needed. From (12) by differentiation

$$\mathfrak{w}_N(x) = NW^{N-1}(x)w(x). \tag{13}$$

The mode will be the solution of

$$\frac{N-1}{W(x)}w(x) + \frac{w'(x)}{w(x)} = 0. \tag{13'}$$

For a given initial distribution $w(x)$ and for small N we have to solve this equation. But the mean and the moments cannot be obtained in a general way by the use of the exact distribution (13). However we can reach general solutions if N is large, provided we limit ourselves to certain classes of initial distributions. We have studied this problem in previous publications [11 13]. For our present purpose it is sufficient to give the results in a form due to R. von Mises [18].

We define a large value u of the variable x by

$$N(1 - W(u)) = 1. \tag{14}$$

This means that the expected number of observations equal to or greater than u is one. Equation (14) is but another form of definition (3). The mean number of trials is used in (3) whereas the original variable x is used in (14).

The probability $\alpha\, du$ that a value greater than u will be contained between u and $u + du$ is given by

$$\alpha = \frac{w(u)}{1 - W(u)}. \tag{15}$$

Obviously α and u are functions of N and the constants in the initial distribution $w(x)$. There are two limiting forms of the probability (12)

$$\lim_{N\to\infty} W^N(x) = F(x); \qquad \lim_{N\to\infty} W^N(x) = \mathfrak{W}(x).$$

If

$$\lim_{u\to\infty} \alpha u = k > 0, \tag{16}$$

we obtain

$$F(x) = e^{-(u/x)^k}. \tag{17}$$

This probability function was first established by Fréchet [5]. If

$$\lim_{u\to\infty} \frac{d}{du}\left(\frac{1}{\alpha}\right) = 0, \tag{18}$$

we obtain

$$\mathfrak{W}(x) = e^{-e^{-\alpha(x-u)}}. \tag{19}$$

This probability function is due to R. A Fisher [4]. Let us consider the first limit. The initial distributions which lead to it belong to the *Pareto type.* For this distribution

$$w(x) = \frac{k}{x^{k+1}}; \qquad W(x) = 1 - \frac{1}{x^k}; \qquad x \geqq 1$$

and condition (16) holds; for *any* value of x

$$\frac{xw(x)}{1 - W(x)} = k.$$

The distribution $f(x)$ of the largest value, which corresponds to (17), is

$$f(x) = \frac{k}{u}\left(\frac{u}{x}\right)^{k+1} e^{-(u/x)^k}. \tag{20}$$

The mode $\tilde{x}_N$ of the largest value is the solution of

$$\frac{d}{dx}\left[(k+1)\ln\frac{u}{x} - \left(\frac{u}{x}\right)^k\right] = 0,$$

hence

$$\frac{k+1}{x} = \frac{ku^k}{x^{k+1}},$$

or

$$\tilde{x}_N = u\left(\frac{k}{k+1}\right)^{1/k}. \tag{21}$$

According to the definition (14) the mode of the largest value will increase with N. For a finite number of observations, which is always the case, the mode will be limited. But the moments of order k or higher will not exist. For $k < 1$, no moment will exist. For $k < 2$, only the first moment, the mean, exists, and so on.

Let us consider now the second limit (19). The initial distributions which lead to it belong to the *exponential type.* For this distribution [14]

$$w(x) = e^{-x}; \qquad W(x) = 1 - e^{-x}; \qquad x \geqq 0,$$

and for *any* value of x

$$\frac{d}{dx}\left(\frac{1 - W(x)}{w(x)}\right) = 0,$$

which means that condition (18) is fulfilled. Most of the distributions used in statistics belong to this type According to (19) the distribution of the largest value is

$$\mathfrak{w}(x) = \alpha e^{-\alpha(x-u)-e^{-\alpha(x-u)}}. \tag{22}$$

If we introduce a reduced variable y without dimension by the linear transformation

$$y = \alpha(x - u), \tag{23}$$

we get the reduced probability $\mathfrak{B}(y)$

$$\begin{aligned} \mathfrak{B}(y) &= \mathfrak{W}(x) \\ &= e^{-e^{-y}}. \end{aligned} \tag{24}$$

The numerical values of this function, calculated by means of Becker's tables [1], are given in Table II, col. 1 and 2. The reduced distribution

$$\mathfrak{v}(y) = e^{-y-e^{-y}}, \tag{25}$$

makes clear the meaning of u. the distribution has one and only one maximum which occurs for the reduced value $y = 0$. Therefore u is the mode of the largest value for a given set of N observations. For an initial distribution $w(x)$ satisfying (18), and for large N, definition (3) of the return period as a function of x becomes identical with relation (14) which involves the number of observations N and the corresponding most probable value u.

We wish to decide which distribution of the largest value is to be used to represent the given observations. This decision depends, according to (16) and (18), on the nature of the initial distribution at the extreme values of the variable. If the law of the observed initial variable is known, a precise answer can be given. But generally speaking, a distribution chosen to represent given observations is nothing but an interpolation formula. Formulas having different analytical properties may all give satisfactory results. One might fulfill condition (16), and another (18). The conditions apply to the differential coefficient, whereas the initial observations are always discontinuous. Therefore they will not enable us to decide which, if any, of the conditions is met. For extreme values of the variable x the observed differences are large and nonuniform, and there is therefore no way to replace the differentiation by a finite difference. Consequently we have to use the observations of the largest values to control the two competing theories and not the conditions. The fact that distribution (20) has higher moments only under certain conditions, is a strong practical argument in favor of distribution (22). Therefore the following development will be based on this distribution.

It can be shown that the mean error θ of distribution (22) is related to the constant α by

$$\theta = 0.98/\alpha. \tag{26}$$

Therefore the constant u is the most probable largest value for N observations and $1/\alpha$ a multiple of the mean error.

TABLE II

Probabilities and return periods of largest values

reduced variable y	probability $\mathfrak{W}(x)$	return period $\log T(x)$	Flood discharges per second	
			in cubic meter x Rhône R.	in 1000 cubic feet x Mississippi R.
−2.00	0.00062	0.000		
−1.75	0.00317	0.001		
−1.50	0.01131	0.005	1355	803
−1.25	0.03049	0.013	1492	869
−1.00	0.06599	0.030	1629	936
−0.75	0.12039	0.056	1766	1002
−0.50	0.19230	0.093	1903	1069
−0.25	0.27693	0.141	2040	1135
0.00	0.36788	0.199	2177	1202
0.25	0.45896	0.267	2314	1268
0.50	0.54524	0.342	2451	1335
0.75	0.62352	0.424	2588	1401
1.00	0.69220	0.512	2725	1468
1.25	0.75088	0.604	2862	1534
1.50	0.80001	0.699	2999	1601
1.75	0.84048	0.797	3136	1667
2.00	0.87342	0.899	3273	1734
2.25	0.89996	1 000	3410	1800
2.50	0.92119	1.103	3547	1867
2.75	0.93807	1.208	3686	1933
3.00	0.95143	1.314	3822	2000
3.25	0.96197	1.420	3959	2066
3.50	0.97025	1.527	4096	2133
3.75	0.97675	1.634	4233	2199
4.00	0.98185	1.741	4370	2266
4.25	0.98584			
4.50	0.98895			
4.75	0.99138			
5.00	0.99329			
5.25	0.99477			
5.50	0.99592			
5.75	0.99682			
6.00	0.99752			

TABLE III

Observed return periods

Rhône, Lyon (France) (1826-1936)

Flood discharge x_m	Serial number m	Return period $\log T(x_m)$	Flood discharge x_m	Serial number m	Return period $\log T(x_m)$
899	1	.004	2475	57	.313
1172	2	.008	2475	58	.321
1231	3	.012	2475	59	.329
1272	4	.016	2491	60	.338
1272	5	.020	2514	61	.346
1432	6	.024	2514	62	.355
1432	7	.028	2514	63	.364
1439	8	.032	2514	64	.373
1444	9	.037	2538	65	.382
1502	10	.041	2554	66	.392
1541	11	.045	2586	67	.402
1560	12	.050	2594	68	.412
1639	13	.054	2594	69	.422
1706	14	.058	2594	70	.432
1780	15	.063	2602	71	.443
1829	16	.068	2626	72	.454
1850	17	.072	2627	73	.465
1857	18	.077	2643	74	.477
1913	19	.081	2675	75	.489
1913	20	.086	2675	76	.501
1934	21	.091	2773	77	.514
1955	22	.096	2773	78	.527
1992	23	.101	2773	79	.540
1992	24	.106	2839	80	.554
2006	25	.111	2856	81	.568
2006	26	.116	2881	82	.583
2013	27	.121	2881	83	.598
2050	28	.126	2965	84	.614
2050	29	.131	3007	85	.630
2072	30	.137	3050	86	.647
2094	31	.142	3058	87	.665
2101	32	.148	3067	88	.684
2115	33	.153	3067	89	.703
2145	34	.159	3126	90	.723
2145	35	.164	3179	91	.744
2153	36	.170	3214	92	.766

TABLE III—*Concluded*

Flood discharge x_m	Serial number m	Return period log $'T(x_m)$	Flood discharge x_m	Serial number m	Return period log $'T(x_m)$
2160	37	.176	3250	93	.790
2168	38	.182	3266	94	.825
2175	39	.188	3293	95	.841
2206	40	.194	3310	96	.869
2206	41	.200	3310	97	.899
2206	42	.206	3354	98	.931
2221	43	.213	3426	99	.966
2236	44	.219	3444	100	1.004
2240	45	.226	3444	101	1.045
2258	46	.232	3480	102	1.091
2281	47	.239	3606	103	1.142
2296	48	.246	3625	104	1.200
2327	49	.253	3708	105	1.267
2342	50	.260	3801	106	1.346
2358	51	.267	3810	107	1.443
2381	52	.274	3905	108	1.568
2420	53	.282	4096	109	1.744
2444	54	.289	4105	110	2.045
2452	55	.297	4390	111	
2467	56	.305			

$\Sigma x_m = 276{,}773.$ $\Sigma x_m^2 = 744{,}538{,}565.$

The arithmetic mean $\bar{u}$ of distribution (22) is [4]

$$\bar{u} = u + \frac{c}{\alpha}, \tag{27}$$

where $c = 0.5772157$ is Euler's constant. The standard deviation s is

$$s = \pi/\alpha\sqrt{6}. \tag{28}$$

Therefore

$$\bar{u} = u + 0.45005s. \tag{29}$$

The reduced variable y introduced by (23) is related to the reduced variable

$$z = \frac{x - \bar{u}}{s} \tag{30}$$

by

$$z = \frac{\alpha\sqrt{6}}{\pi}(x - u) - \frac{c\sqrt{6}}{\pi}.$$

The substitution of the numerical values leads to

$$z = 0.77970y - 0.45005. \tag{30'}$$

Conversely,

$$y = 1.28255z + 0.57722. \tag{31}$$

The value (32) $v = s/\bar{u}$, the coefficient of variation, is related to the product αu. By (27) $\alpha u = \alpha\bar{u} - c$ and by (28)

$$\alpha u = \frac{\pi}{\sqrt{6}} \cdot \frac{1}{v} - c. \tag{33}$$

Therefore the numerical value of αu can also be considered as a characteristic of an observed distribution of largest values.

For the two constants we calculate for the observed distribution of largest values the two first moments

$$\bar{u} = \frac{1}{n}\sum_{m=1}^{n} x_m, \tag{34}$$

and

$$\overline{u^2} = \frac{1}{n}\sum_{m=1}^{n} x_m^2. \tag{35}$$

To get the observed standard deviation we use the Gaussian formula

$$s = \sqrt{\left(1 + \frac{1}{n-1}\right)(\overline{u^2} - \bar{u}^2)}. \tag{36}$$

According to (28) and (27)

$$\frac{1}{\alpha} = 0.7796968s, \tag{37}$$

and

$$u = \bar{u} - \frac{0.5772157}{\alpha}. \tag{38}$$

These formulas give the two constants in the distribution of largest values.

3. Flood flows interpreted as largest values. We will now apply the theory of largest values to flood flows. Let us consider the daily flow as a statistical variable, unlimited to the right. This idea is not new. The formulas proposed by Fuller [7], Hazen [15], and numerous other authors all incorporate this assumption. Gibrat [9] supposes that the daily flows vary according to Galton's distribution. Instead of postulating a specific formula for the distribution of flows we shall only suppose that it belongs to the usual exponential type, which means that condition (18) is fulfilled.

We define a flood as being the largest value of the $N = 365$ daily flows. The

flood flows are therefore the largest values of flows. This commonplace implies the distinction between floods and inundations. For each year there exists one or more floods of the same magnitude, but there might exist several different inundations or none at all. If there are several inundations in a year the greatest one will be a flood; but a flood need not to be an inundation even a dry year has a flood. We limit ourselves to floods, assume that $N = 365$ is a large number, and represent the distribution of annual floods by the distribution (22) of largest values.

There have been objections to the concept that the daily flow is an unlimited variable. Horton [16] believes that this implies the absurd idea of unlimited floods. This opinion is shared by Slade [20], who claims that there is a definite upper limit to the magnitude of the floods for a given stream. The theory of largest values confirms only partially Horton's opinion. If we should choose distribution (20), the most probable annual flood will be limited For this distribution, however, it might happen that the mean annual flood has no meaning. To avoid this we have chosen distribution (22), for which the mean annual flood and all the moments will be finite. A further justification of the use of (22) might be derived from the fact that Galton's distribution belongs to the exponential type. As a final argument, numerical calculations show that formula (22) gives a better fit to the observed distributions of flows.

The variable x is the annual flood flow measured in cubic meters or cubic feet per second. The mean $\bar{u}$ is the annual mean flood, whereas u is the most probable annual flood. The value s is the standard deviation of the distribution of annual floods. Finally y is called the reduced flood.

The distribution (22) possesses the properties of the observed distribution of flood flows. It is asymmetrical; rising rather quickly but falling rather slowly. The modal value is to the left of the mean (see Fig 3).

To apply the theory of return periods let us consider the event of the highest annual discharge being greater than x. We have to replace in formula (3) the general probability $W(x)$ by the probability of flood discharges (19). The number of observations n is the number of years for which observations exist.

To use formula (3) we have to suppose that the intervals between the successive floods are all equal to one year. This assumption conforms more or less to the seasonal nature of floods.

The return period of a flood greater than x

$$T(x) = \frac{1}{1 - e^{-e^{-y}}} \tag{39}$$

is the arithmetic mean of the intervals between two years, which have a flood discharge greater than x; the discharges for the intervening years are all less than x. Therefore $T(x)$ is the mean of the number of years for which x will be surpassed once Formula (39) gives the meaning of u from the standpoint of the return period. For $y = 0$

$$T(u) = \frac{e}{e - 1}.$$

The return period $T(u)$ of the most probable annual flood is 1.58198 years. In other words, the constant u is the flood discharge with return period

$$\log T(u) = 0.19920 \tag{40}$$

where log signifies the common logarithm. The return period of the mean annual flood is by (27) and (39) equal to 2.32762 years.

Let us now consider the relation between the flood discharge x and its return period for small and large values of x. To small values of x correspond large negative values of y and therefore return periods T approximating 1. The distribution (25) of the largest values being unlimited, the flood discharge considered as a function of log T will by (6) increase rapidly at first. To large values of x correspond large values of y and $T(x)$. If we introduce the natural logarithm, (39) gives

$$-\ln\left(1 - \frac{1}{T(x)}\right) = e^{-y}.$$

For large values of x, viz., $T(x) \geqq 10$, it is sufficiently accurate to use

$$\frac{1}{T(x)} = e^{-y},$$

so that

$$y = \ln T(x). \tag{41}$$

If the common logarithm is used,

$$\log T(x) = 0.434294\alpha(x - u). \tag{42}$$

The logarithm of the mean number of years for which the flood discharge will once be exceeded, converges towards a linear function of x. This property of the distribution of largest values was established by M. Coutagne [2]. Let us write

$$x = u + \frac{2.30258}{\alpha} \log T(x). \tag{43}$$

Then $1/\alpha$ can be considered as a measure of the increase of a flood discharge with respect to the logarithm of time

According to the general formulas (6) and (42) the shape of the return period as a function of the flood discharge x is as follows: at the beginning i.e., for small flood discharge, the return periods are close to 1 and increase very slowly. At the end, i.e., for large flood discharges, the logarithm of the return period converges to a linear function of x.

Another form of (43) is

$$\frac{x}{u} = 1 + \frac{2.30258}{\alpha u} \log T(x). \tag{44}$$

The ratio of the flood discharge which will be exceeded in the mean once in T years to the modal annual flood converges to a linear function of the logarithm

of the return period. The constant $1/\alpha u$ of dimension zero depends, by (33), on the coefficient of variation. Its value is a characteristic of the stream. If we introduce the arithmetic mean $\bar{u}$ and the standard deviation s we obtain by (42), (27), and (28)

$$x = \bar{u} - 0.45005s + (0\ 77970)\ (2.30258)s \log T(x).$$

Therefore, approximately,

$$\frac{x}{\bar{u}} = 1 - \frac{9}{20}v + 1.796v \log T(x). \tag{45}$$

The right hand member of this linear equation contains only one constant, the coefficient of variation of the floods Finally by (42) and (31)

$$\log T(x) = 0.25068 + 0.55700\frac{x - \bar{u}}{s}. \tag{46}$$

There is still another way of interpreting these asymptotic formulas. Let $T(2x)$ be the return period of the value $2x$, then by (43)

$$2x = u + \frac{\ln T(2x)}{\alpha},$$

therefore

$$2 = \frac{\alpha u + \ln T(2x)}{\alpha u + \ln T(x)},$$

and finally

$$T(2x) = T^2(x)e^{\alpha u}. \tag{47}$$

The return period of a flood of magnitude $2x$ is equal to the square of the return period of x multiplied by a factor which depends only upon the coefficient of variation.

All these asymptotic formulas are good approximations only for return periods above ten years, which means according to Table II, $y \geqq 2.25$ or according to (23), (30) and (31) $x \geqq \bar{u} + 1.3s$. The corresponding value of the flood probability is by (3) $\mathfrak{W}(x) \geqq 0.9$. The consequences of (41) can be applied to only 10% of the observations, i.e. to the large flood discharges. Their observed return periods are based on a few observations and may therefore differ considerably from the theoretical values. In spite of the above restrictions the linear formula (43) has a meaning for values of T equal to or greater than unity. We now ask: How will the most probable largest value increase with the number of observations? This number of years can again be called T. The answer to the above question requires the solution of (13′) where the distribution (25) of largest values $\mathfrak{v}(y)$ must be introduced as the initial distribution $w(x)$.
From (24)

$$\frac{T - 1}{e^{-e^{-y}}}\, e^{-y-e^{-y}} - 1 + e^{-y} = 0,$$

or

$$Te^{-y} = 1,$$

which is identical with (41). For $T = 1$ the most probable annual flood is of course u. Therefore the relation (41), valid for $T \geqq 1$, means: The most probable flood $u(T)$ to be reached within T years is a linear function of the logarithm of T

$$u(T) = u + \frac{2.30258 \log T}{\alpha}. \tag{41'}$$

The constant $1/\alpha$ is the slope of this straight line. The results (41–46) are related to Fuller's well-known formula [6]. This author, the first to investigate flood flows systematically, proposed a linear relation between the logarithm of the return period and the arithmetic mean of the flood discharges greater than the mth value (m taken from above). A similar empirical formula has been stated by Lane [7] and has been applied by Saville [19]. The similarities and differences between these interpolation formulas and our theory can be stated in the following way: If we start from the theory of largest values we reach these formulas as asymptotic expressions for the return period of large floods. Considered this way, our theory gives a certain justification to Fuller's hypothesis. But Fuller's and similar formulas were intended to apply to all flood discharges. Now, the distribution of the flood discharges (4) corresponding to these return periods does not fit the observations. It can be shown that these formulas involve the assumption of a simple exponential distribution $\varphi(x)$ for the flood discharges

$$\varphi(x) = \frac{1}{\bar{u} - \epsilon} e^{-(x-\epsilon)/(\bar{u}-\epsilon)}; \tag{48}$$

and the existence of a lower limit ϵ of the flood discharges given by $\epsilon = \bar{u} - s$. In Fuller's formula all flood discharges must be greater than 2/3 of the mean annual flood. The density of probability always diminishes with increasing magnitude of the flood This neglects the ascending branch (about one third) of the distribution of floods (see Fig. 3) and is incompatible with the observed facts. We therefore prefer our formula which takes account of the total variation, but we do not minimize the importance of Fuller's work which has led to much valuable research

Formula (39) gives the theoretical return periods $T(x)$ as a function of the reduced flood discharge y, and holds for the entire range of observations. The general numerical values are given in Table II, cols. 1 and 3. For a given stream, the return period of a flood discharge greater than x depends by (23) upon the two constants α and u. If these values have been calculated by (37) and (38) the theoretical flood discharge x corresponding to $T(x)$ is obtained by the linear transformation

$$x = u + y/\alpha. \tag{49}$$

The asymptotic formula (42) suggests the coordination of the flood discharges x and the logarithm of the return periods.

4. Rhône and Mississippi Rivers. We think that our system of formulas is simple, logically consistent and free of artificial assumptions. Now it remains to be shown that the arithmetic involved is simple and that the results fit the observations For the Rhône we shall analyze the observed cumulative frequency, the distribution, and the return periods. For the Mississippi River we shall limit ourselves to the return periods.

For each year we choose the maximum of the daily discharges (we do not use momentary peaks). The 111 values x_m for the Rhône 1826–1936 published by Coutagne [3] and arranged in order of increasing magnitude are given in Table III (col. 1). The supposition that the intervals between consecutive floods are all equal to one year is not always true. Only 77 of the 111 floods occurred between October and March, whereas 34 were scattered throughout the year. But the

TABLE IV

Calculation of constants

Stream observation station	Rhône Lyon (France) 1826–1936	Mississippi River Vicksburg (Miss.) 1890–1939
Number of observations n	111	50
Annual mean flood $\bar{u}$	2,493.5	1,355.6
Mean squared flood $\overline{u^2}$	6,707,555.0	1,951,828.8
Standard deviation s	703.1	341.3
Constant $1/\alpha$	548.2	266.1
Most probable annual flood u	2,177.0	1,201 9

differences in the lengths of the intervals compensate each other. The second column of Table III contains the serial number m According to (9) we calculate for the mth observed flood discharge x_m, taken in ascending magnitude, the logarithm of the observed return period $\log n/(n-m)$ (col. 3), where $n = 111$ and $m = 1, 2, \cdots, 110$, and obtain the exceedance intervals. The other observed curve, the recurrence interval, is obtained by (10) through the coordination of x_{m+1} and $\log n/(n-m)$. Both curves are plotted in Fig. 1. The recurrence and exceedance intervals differ for the large flood discharges. The observed flood discharges arranged in increasing magnitude are plotted in the cumulative histogram, Fig. 2.

To compare these observations with our theory, we calculate the two constants $1/\alpha$ and u according to the formulas (34)–(38). The values Σx_m and Σx_m^2 are given at the end of Table III. Division by $n = 111$ gives the mean flood $\bar{u}$ and the mean squared flood $\overline{u^2}$ (Table IV). The Gaussian correction being $1 + 1/110$ we obtain from formula (36) the standard deviation s (Table IV)

TABLE V

Observed and theoretical distributions of flood discharges

Rhône

Reduced variable y	Variable x	Midpoints $x + \frac{\Delta x}{2}$	Observed distribution $111\Delta' \mathfrak{W}(x)$	Theoretical distribution $111\Delta \mathfrak{W}(x)$	Cumulative frequency $111 \mathfrak{W}(x)$
−2.75	670				
−2.50		807	1		0.00
−2.25	944			0.01	0.01
−2.00		1081	1	0.34	0.07
−1.75	1218			1.19	0.35
−1.50		1355	7	3.03	1.26
−1.25	1492			6.07	3.38
−1.00		1629	5	9.98	7.33
−0.75	1766			14.02	13.36
−0.50		1903	13	17.38	21.35
−0.25	2040			19.49	30.74
0.00		2177	21	20.21	40.84
0.25	2314			19.68	50.95
0.50		2451	19	18.26	60.52
0.75	2588			16.31	69.21
1.00		2725	14	14.14	76.83
1.25	2862			11.97	83.35
1.50		2999	9	9.94	88.80
1.75	3136			8.15	93.29
2.00		3273	8	6.61	96.95
2 25	3410			5.30	99.90
2.50		3547	6	4.23	102.25
2.75	3686			3.45	104.13
3.00		3822	4	2.65	105.70
3.25	3959			2.00	106.78
3.50		4096	2	1.64	107.70
3.75	4233			1.28	108.42
4.00		4370	1	1.01	108.98
4.25	4507			0.79	109.43
4.50		4644	0	0.61	109.77
4.75	4781			0.48	110.04
5.00		4918		0.38	110.25
5.25	5055			0.30	110.42
5 50		5192		0.23	110.55
5.75	5329			0.18	110 65
6.00		5466		0.27	110.73
			111	111.00	

and finally from (37) and (38) the constant $1/\alpha$ and the most probable annual flood u From the numerical values in Table IV the linear transformation (49) for the Rhône is

$$x = 2177.03 + 548\ 19y$$

TABLE VI

Observed return periods

Mississippi River, Vicksburg, (Miss.) (1890–1939)

Flood discharge x_m	Serial number m	Return period $\log' T(x_m)$	Flood discharge x_m	Serial number m	Return period $\log' T(x_m)$
760	1	0 0088	1357	26	3188
866	2	.0178	1457	27	.3273
870	3	0269	1397	28	3566
912	4	.0362	1397	29	.3768
923	5	.0458	1402	30	.3980
945	6	0555	1406	31	.4202
990	7	0655	1410	32	.4437
994	8	.0758	1410	33	.4686
1018	9	0862	1426	34	.4949
1021	10	.0969	1453	35	.5229
1043	11	1079	1475	36	.5529
1057	12	.1192	1480	37	.5851
1060	13	1308	1516	38	.6198
1073	14	.1427	1516	39	.6576
1185	15	.1549	1536	40	.6990
1190	16	.1675	1578	41	.7448
1194	17	.1805	1681	42	.7959
1212	18	.1939	1721	43	8539
1230	19	.2076	1813	44	9208
1260	20	2219	1822	45	1 0000
1285	21	.2366	1893	46	1.0969
1305	22	2518	1893	47	1.2219
1332	23	.2676	2040	48	1.3980
1342	24	.2840	2056	49	1.6990
1353	25	.3011	2334	50	

$\Sigma x_m = 67{,}780.$ $\Sigma x_m^2 = 97{,}591{,}440.$

This leads to the determination of the theoretical flood discharges. The theoretical return periods $\log T(x)$ are given in Table II, col. 3 as a function of the reduced variable y and of x (col. 4). The discharges x obtained by letting y take on the values -2.75 to 6.00 in the linear transformation, are given in

Table V, cols. 2 and 3 and plotted in Fig. 1. The distances Δx used in the calculations of the theoretical discharges are $1/4\alpha = 137.05$.

Along the abscissa are plotted the logarithm of the return periods and the return periods in years; along the ordinate are plotted the corresponding flood discharges and the modal annual flood u. The straight line from the point $(u, 0)$ to the asymptote gives the most probable flood as a function of time. The theoretical curve corresponds quite closely with the general course of the observations. For small floods the theoretical return periods are practically iden-

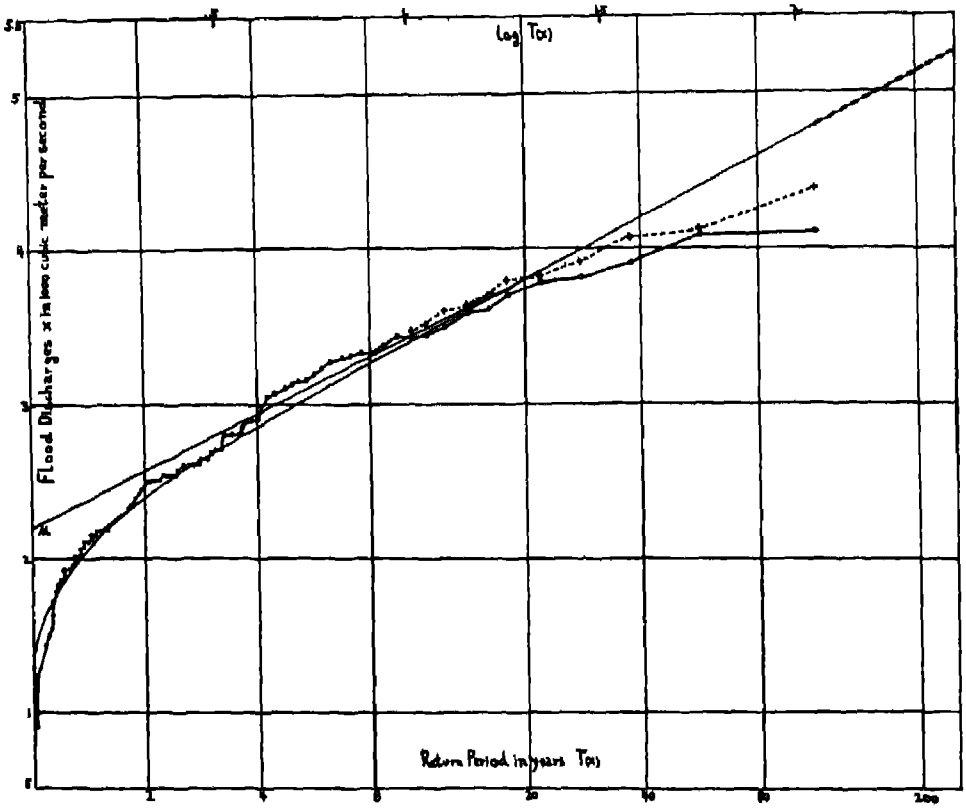


FIG. 1. RHÔNE AT LYON (FRANCE) 1826–1936

Observations Table III: Recurrence intervals, + – – +; Exceedance intervals, •——•; Return periods, ———; Theory Table II, cols. 3 and 4: Extrapolation, — —.

tical with the observed values. But for the very large floods the theoretical curve surpassed both the exceedance and recurrence intervals.

The observed cumulative histogram is shown in Fig. 2. We calculate from Table II, col. 2, the frequencies $111\mathfrak{W}(x)$ (Table V, col. 6). These theoretical values $(x, 111\mathfrak{W}(x))$ are also plotted in Fig. 2. The agreement between theory and observations is very good.

For the comparison of the observed and theoretical distributions of the flood discharges we use what might be called the natural classification. For the

observations, the length of the class intervals and the beginning of the first class interval are arbitrary. In order to obtain the observed distribution of the flood discharges, it is natural to use the theoretical class intervals set forth in Table V, col. 2. The data of the third column can be interpreted as the midpoints of the class intervals given in col. 2. The frequencies for these class intervals are ob-

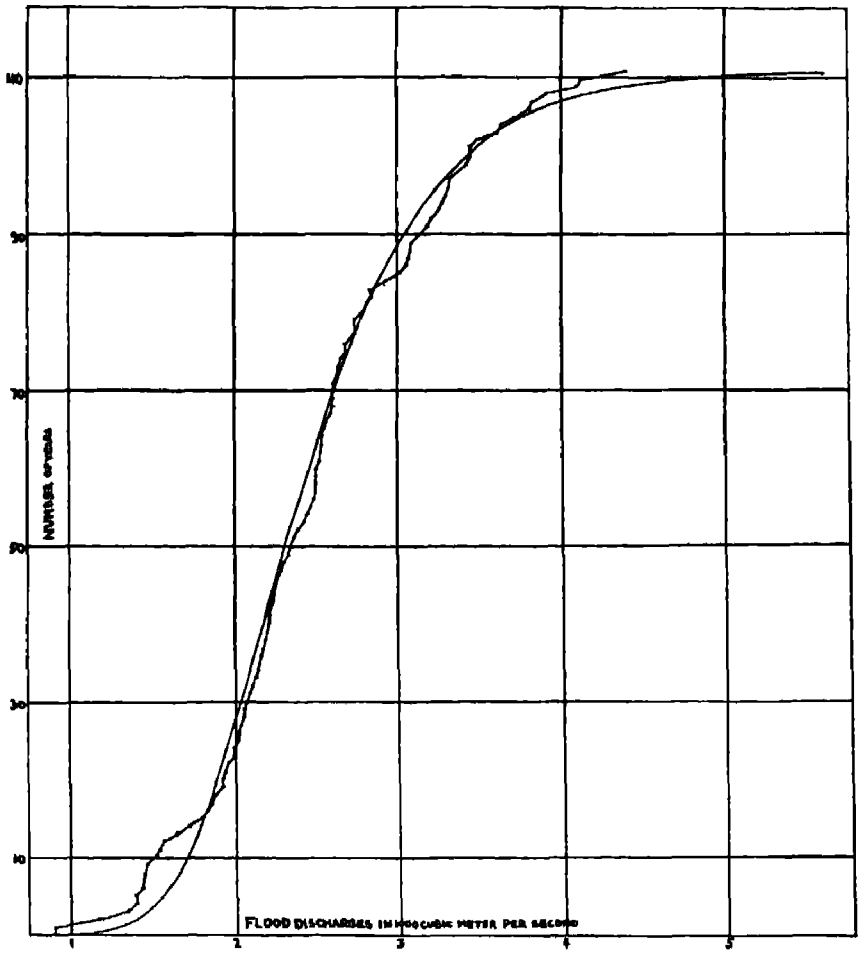


FIG 2 CUMULATIVE FREQUENCY OF THE FLOOD DISCHARGES. RHÔNE, LYON (FRANCE) 1826–1936

Observations Table III cols 1 and 2, •—•; Theory Table V cols. 2, 3 and 6, /

tained from Table III, and are given in Table V, col. 4. The observed distribution is shown in Fig. 3. To obtain the corresponding theoretical distribution we calculate from Table V, col. 6, the difference between two cumulative frequencies disjoined by one, i.e., we pair consecutively the first and third, the second and fourth items and so on This theoretical distribution given in col. 5 and the observed distribution are based on class intervals of the same length. Fig. 3

shows that the theoretical distribution $\Delta\mathfrak{W}(x)$ of the largest values agrees in a satisfactory way with the observed distribution $\Delta'\mathfrak{W}(x)$ of the flood discharges. Table VI, col. 1, gives the corrected[2] flood discharges x_m, measured in units of 1000 cubic feet per second, for the Mississippi River at Vicksburg (1890–1939), ($n = 50$), arranged according to increasing magnitude; col. 2 gives the serial number m. We calculate the logarithm of the observed return periods $\log n/(n-m)$, (col. 3). The observations $(x_m, \log {}'T(x_m))$ and $(x_{m+1}, \log {}'T(x_m))$ are plotted in Fig. 4. The constants obtained by formulas (34)- (38) are shown

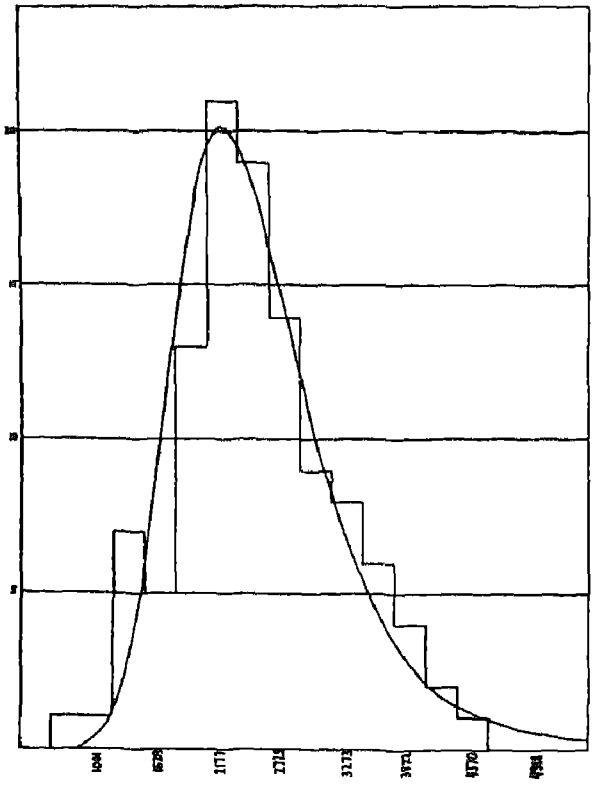


FIG. 3. DISTRIBUTION OF THE FLOOD DISCHARGES RHÔNE, LYON (FRANCE) 1826–1936
Observations Table V cols. 2, 3 and 4, ⊓; Theory Table V cols 2, 3 and 5, ⌒

in Table IV. By (49) the theoretical floods x corresponding to the return periods $T(x)$ presented in Table II, col. 3, are

$$x = 1201.98 + 266.14y.$$

These floods are given in Table II, col. 5. The class interval used is

$$1/4\alpha = 66.5.$$

[2] These data have been put at my disposal through the courtesy of Mr. A. E. Brandt of the U. S. Department of Agriculture

The theoretical curve $(x, \log T(x))$, plotted in Fig. 4, agrees in a very satisfactory way with the observations. For the large floods the theoretical return periods are between the exceedance and recurrence intervals.

The calculations of the theoretical return periods for other streams, e.g. the Columbia, Connecticut, Cumberland, Rhine, and Tennessee Rivers, for which reliable observations exist for more than 60 years, also show a good agreement with the observations The goodness of fit diminishes for streams for which the number of observations is smaller and for which the data are not very reliable.

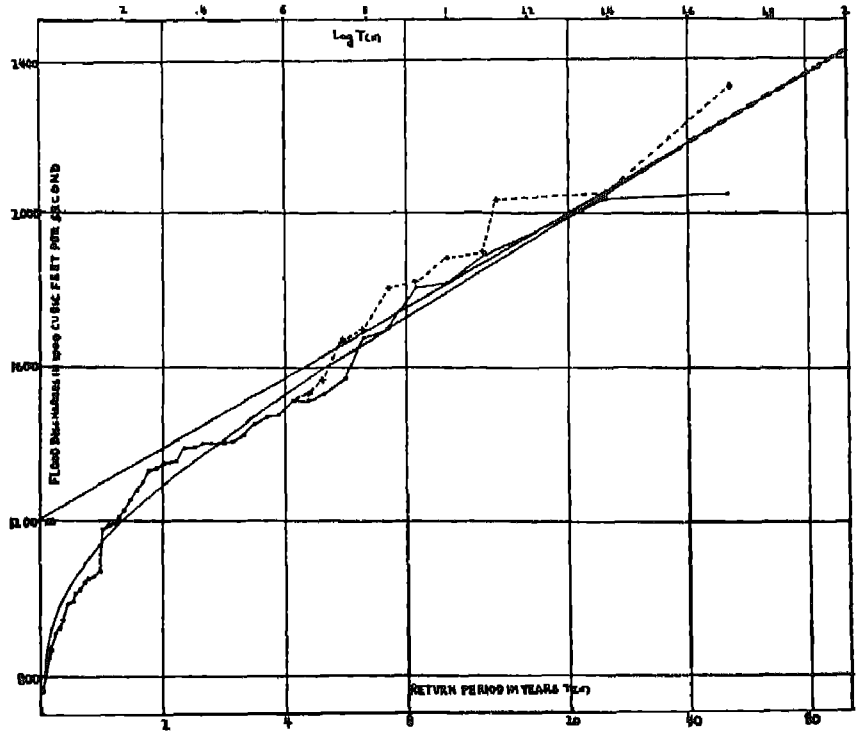


FIG. 4. MISSISSIPPI RIVER AT VICKSBURG, (MISS) 1890–1939

Observations Table VI Recurrence intervals, + – – +, Exceedance intervals, •——•, Return periods, ———, Theory Table II, cols. 3 and 5, Extrapolation, — —.

5. Summary and conclusions. In order to apply any theory we have to suppose that the data are homogeneous, i.e. that no systematical change of climate and no important change in the basin have occurred within the observation period and that no such changes will take place in the period for which extrapolations are made. It is only under these obvious conditions that forecasts can be made.

The theoretical return period $T(x)$, the mean number of years between two annual flood discharges greater than or equal to x, is a statistical function such as the distribution $w(x)$ or the probabilities $W(x)$ and $P(x)$. There are two

sets of observed values corresponding to the theoretical set. The exceedance interval $'T(x_m)$ formula (9), and the recurrence interval $''T(x_m)$ formula (9'); x_m being the mth flood discharge, where m is counted from below. As any theory must include both notions, no separate theory for exceedance or recurrence intervals is possible.

The return period $T(x)$ of a flood discharge x is found by formula (39). For large values of x the flood discharge converges toward a linear function (42) of the logarithm of the return period. This is the scientific basis of Fuller's empirical formula. The two constants of our formula u and $1/\alpha$, are, respectively, the most probable annual flood discharge and a multiple of the standard deviation (28). Their values depend upon the drainage basin and known geological and meteorological factors. It is beyond our present task to consider the influence of these factors. Our method can be summarized by the following rules:

1) For each year find the maximum daily discharge x_m (do not use momentary peaks) and arrange these n data in increasing magnitudes.

2) Calculate for each discharge x_m ($m = 1, 2, \cdots, n-1$), the values $\log {}'T(x_m) = \log n - \log (n-m)$ and plot the curves x_m, $\log n/(n-m)$, and x_{m+1}, $\log n/(n-m)$. These are the observed exceedance and recurrence intervals.

3) Calculate the annual mean flood $\bar{u}$ and the annual mean squared flood $\overline{u^2}$; determine according to (36)–(38) the standard deviation

$$s = \sqrt{\left(1 + \frac{1}{n-1}\right)(\overline{u^2} - \bar{u}^2)}$$

and the two constants

$$1/\alpha = 0.77970s,$$

$$u = \bar{u} - \frac{0.57722}{\alpha}.$$

4) The theoretical flood discharges x corresponding to the logarithm of the return period $T(x)$ given in Table II, col. 3, are obtained by the linear transformation

$$x = u + y/\alpha$$

where y is taken from Table II, col. 1. Plot x as a function of $\log T(x)$. For large values of x and for extrapolation it is sufficient to use the linear asymptote obtained graphically.

The linear part of the theoretical curve (x, $\log T$) permits of two interpretations: First, T is the theoretical return period of a flood greater than or equal to x; second, x is the most probable flood to be reached within T years. The second interpretation holds for the straight line through the point (u, 0).

The figures show a close agreement between observed and theoretical values.

The observed curvature of the return periods is brought out by the theoretical graph.

The agreement between theory and observation is excellent for floods which correspond to reduced values of $y \leqq 3$. For the two or three extreme floods, the return periods are based on a few observations and, consequently, the agreement is not very good. No theory can be verified by two or three observations. Generally speaking, the theory fits the observations as closely as could be expected for such a complicated phenomenon

In order to make a further test of our results, we need a numerical measure for the weights to be given to the theoretical points Therefore, for a given probability we must find the corresponding theoretical limits for the observed return periods. The theory of positional values will give these control curves. Since it was the purpose of this article to develop and make clear the basic method, we have refrained from introducing this subject.

It is our claim that the calculus of probabilities and especially the theory of largest values, is an efficient tool for the solution of certain hydrological problems.

REFERENCES

[1] G. E. Becker and C. E. van Orstrand, *Hyperbolic Functions*, Smithsonian Mathematical Tables, Washington, 1931

[2] A. Coutagne, "Etude statistique des débits de crue," *Revue Générale de l'Hydraulique* Paris (1937).

[3] A Coutagne, "Etude statistique et analytique des crues du Rhône à Lyon," *Comptes Rendus du Congrès pour l'Utilisation des Eaux*, Lyon, (1938).

[4] R. A Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the smallest and the largest member of a sample," *Proc Camb. Phil. Soc.*, Vol. 24 (1928)

[5] M. Fréchet, "Sur la loi de probabilité de l'écart maximum," *Annales Soc Polon. Math.*, Vol. 6, (1927)

[6] Weston E Fuller, "Flood flows," *Trans. Am Soc. Civil Eng.*, Vol. 77 (1914).

[7] Weston E Fuller, E. Lane and others, "Discussion on flood flow characteristics," *Trans. Am. Soc Civil Eng.*, Vol. 89 (1926).

[8] John C. Geyer, "New curve fitting method for analysis of flood-records," *Trans. Am Geophy. Union*, Part II (1940), pp. 660–668

[9] Robert Gibrat, "Aménagement hydro-électrique des cours d'eau," "Statistique mathématique et calcul des probabilités, "*Revue Générale de l'Electricité*, Vol. 32, No. 15, 16, Paris (1932).

[10] Eugene L. Grant, "The probability-viewpoint in hydrology," *Trans. Am Geophy. Union*, Part 1 (1940), pp. 7–12.

[11] E. J. Gumbel, "Les valeurs extrêmes des distributions statistiques," *Annales de l'Institut Henri Poincare*, Vol 4 (1935), p 115.

[12] E. J. Gumbel, "La plus grande valeur," *Aktuarske Vedy*, Vol. 5, No. 2, p. 83, No. 3, p. 133, No. 4, p. 146, Prague (1935–36).

[13] E. J. Gumbel, *La Durée Extrême de la Vie Humaine, Actualités Scientifiques et Industrielles*, Hermann et Cie, Paris, 1937

[14] E. J. Gumbel, "Les intervalles extrêmes entre les émissions radioactives," *Jour. de Phys.*, Serie 7, Vol. 8, No. 8, No 11 (1937).

[15] Allen Hazen, *Flood Flows, A Study of Frequencies and Magnitudes*, John Wiley and Sons, Inc., New York, 1930.

[16] ROBERT E HORTON, "Hydrologic conditions as affecting the results of the application of methods of frequency analysis to flood records," *Geological Survey Water-Supply Paper* 771, Washington (1936)

[17] CLARENCE S. JARVIS, "Floods in the United States, Magnitude and Frequency," *Geological Survey Water Supply Paper* 771, Washington (1936).

[18] R. VON MISES, "La distribution de la plus grande de n valeurs," *Revue Math. de l'Union Interbalkanique*, Vol. 1, Athens (1936).

[19] T. SAVILLE, "A study of methods of estimating flood flows applied to the Tennessee River," *Publications from College of Engineering*, Nr 6, New York (1935 36).

[20] J. J SLADE, "The reliability of statistical methods in the determination of flood frequencies," *Geological Survey Water-Supply Paper* 771, Washington (1936)

ON THE FOUNDATIONS OF PROBABILITY AND STATISTICS[1]

By R. von Mises

Harvard University

1. Introduction. The theory of probability and statistics which I have been upholding for more than twenty years originates in the conception that the only aim of such a theory is to give a description of certain observable phenomena, the so called mass phenomena and repetitive events, like games of chance or some specified attributes occurring in a large population. Describing means here, in the first place, to find out the relations which exist between sequences of events connected in some way, e.g. a sequence of single games and the sequence composed of sets of those games or between a sequence of direct observations and the so called inverse probability within the same field of observations. The theory is a mathematical one, like the mathematical theory of electricity, based on experience, but operating by means of mathematical processes, particularly the methods of analysis of real variables and theory of sets.

We all know very well that in colloquial language the term probability or probable is very often used in cases which have nothing to do with mass phenomena or repetitive events. But I decline positively to apply the mathematical theory to questions like this. What is the probability that Napoleon was a historical person rather than a solar myth? This question deals with an isolated fact which in no way can be considered as an element in a sequence of uniform repeated observations. We are all familiar with the fact that, e.g. the word energy is often used in every day language in a sense which does not conform to the notion of energy as adopted in mathematical physics. This does not impair the value of the precise definition of energy used in physics and on the other hand this definition is not intended to cover the entire field of daily application of the term energy.

We discard likewise the scholastic point of view displayed in a sentence of this kind: ". . . that both in its meaning and in the laws which it obeys, probability derives directly from intuition and is prior to objective experience." This sentence is quoted from a mathematical paper printed in a mathematical journal of 1940. The same author continues calling probability a metaphysical problem and speaking of the difficulties "which must in the nature of things always be encountered when an attempt is made to give a mathematical or physical solution to a metaphysical problem." In my opinion the calculus of probability has nothing to do with metaphysics, at any rate not more than geometry or mechanics has.

[1] Address delivered on September 11, 1940 at a meeting of the Institute of Mathematical Statistics in Hanover, N. H

On the other hand we claim that our theory, which serves to describe observable facts, satisfies all reasonable requirements of logical consistency and is free from contradictions and obscurities of any kind. I am now going to outline the essential ideas of the theory as developed by me since 1919 and I shall have to refer as to the proof of its consistency to the recent work of A. H. Copeland, of J. Herzberg and of A Wald. Then I will give some examples of application in order to show how the theory works and how it applies to actual problems in statistics.

2. The notion of kollektiv. The basic notion upon which the theory is established is the concept of *kollektiv*. We consider an infinite sequence of experiments or observations every one of which supplies a definite result in the form of a number (or a group of numbers in the case of a kollektiv of more than one dimension). We shall designate briefly by X the sequence of results x_1, x_2, x_3, $\cdots$. In tossing a die we get for X an endless repetition of the integers one to six, $x = 1, 2, \cdots 6$ If we are interested in death probability, we observe a large group of healthy 40 year old men and mark a one for each individual surviving his 41st aniversary and a zero for each man who dies before, so that the sequence x_1, x_2, x_3, $\cdots$ consists of zeros and ones. In a certain sense the kollektiv corresponds to what is called a *population* in practical statistics. Experience shows that in such sequences the relative frequency of the different results (one to six in the first of our examples, one and zero in the second) varies only slightly, if the number of experiments is large enough. We are therefore prompted to assume that in the kollektiv, i.e. in the theoretical model of the empirical sequences or populations, each frequency has a *limiting value*, if the number of elements increases endlessly. This limiting value of frequency is called, under certain conditions which I shall explain later, the "probability of the attribute in question within the kollektiv involved." The set of all limiting frequencies within one kollektiv is called its *distribution*.

Let me insist on the fact that in no case is a probability value attached to a single event by itself, but only to an event as much as it is the element of a well defined sequence. It happens often that one and the same fact can be considered as an element of different kollektivs. It may then be that different probability values can be ascribed to the same event. I shall give a striking example of this, which we encounter in the field of actual statistical problems, at the end of this lecture.

The objection has been made: Since all empirical sequences are obviously finite sequences, why then assume infinite kollektivs? Our answer is that any straight line we encounter in reality has finite length, but geometry is based on the notion of infinite straight lines and uses e.g. the notion of parallels which has no sense, if we restrict ourselves to segments of finite lengths. Another objection, often repeated, reads that there is a contradiction between the existence of a frequency limit and the so called Bernoulli theorem which states that sequences of any length showing a frequency say $\frac{1}{4}$ can also occur in cases for

which the probability equals $\frac{1}{2}$. But it has been proved, in a rigorous way excluding any doubt, that the two statements are compatible, even by explicit construction of infinite sequences fulfilling both conditions. I would even claim that the real meaning of the Bernoulli theorem is inaccessible to any probability theory that does not start with the frequency definition of probability.

Now we are in the position to explain how our probability theory works. This sequence of zeros and ones

$$(X) \quad 1\,0\,1\,|\,0\,0\,1\,|\,1\,0\,0\,|\,0\,1\,1\,|\,1\,1\,0\,|\,0\,1\,1\,|\,0\,1\,0\,|\,1\,1\,1\,\cdots$$

may represent the outcomes of a game of chance The ones show gains, the zeros losses for one of the two players If we separate the terms of X into groups of three digits and replace each group by a single one or zero according to the majority of terms within the group, we get a new sequence

$$(X') \qquad 1\,0\,0\,1\,1\,1\,0\,1\,\cdots$$

which represents the gains and losses in sets of three games Our task is now to compute the distribution, i.e. the limiting frequencies of zeros and ones in this new sequence X', assuming the two frequencies in X are known A sequence can formally be considered as a unique number like a decimal fraction with an infinite number of digits. Then the transition from X to X' can be called a *transformation of a number* $X' = T(X)$. As our sequences have to fulfill certain conditions Copeland calls the sequences X, X' admissible numbers. What I just quoted was of course a very special example of a transformation of a number. But we have to emphasize that all problems dealt with in probability theory, without any exception, have this unique form: The distribution or the limiting frequencies in certain sequences are given, other sequences are derived from the given ones by certain operations, and the distributions in these derived sequences have to be computed. In other words: *Probability theory is the study of transformations of admissible numbers, particularly the study of the change of distributions implied by such transformations*

We know four and only four simple, i.e. irreducible transformations or *four fundamental operations*. They are called selection, mixing, partitioning and combination. By combining these basic processes we can settle all problems in probability theory. The formal, mathematical difficulties in carrying out the computation of the new distributions may become very serious in certain cases, particularly if we have to apply an infinite number of transformations (asymptotic problems). But, in the clearly defined framework of this theory no space is left for any metaphysical speculations, for ideas about sufficient reason or insufficient reason, for notions like degree of evidence or for a special kind of probability logic and so on. And further no modification is needed for handling usual statistical problems: Terms like inverse probability, likelihood, confidence degrees, etc. are justified and admitted only as far as they are capable of being reduced to the basic notion of kollektiv and distribution within a kollektiv. I will give some more details to this point later. Meanwhile let me turn to a

general question which, in a certain way, is the crucial point in establishing the new probability theory.

3. Place selections and randomness. It is obvious that we have to restrict still further the notion of kollektiv or the field of sequences which can be considered as the objects of a probability investigation. The successive outcomes of a game of chance differ very clearly from any regular sequence as defined by a simple arithmetical law, e.g. the regularly alternating sequence 0 1 0 1 0 1 0 1 $\cdots$. A typical property which singles out the irregular or random sequences and which has to be reproduced in every probability theory is that, if p is the probability of encountering a one in the sequence, then p^2 is the probability of two ones following each other immediately. Any probability theory has to introduce an axiom which enables us to deduce this theorem and others of a similar type The question is only how to find a sufficiently general and consistent form for it. The procedure I have chosen consists in using a special kind of transformation of a sequence, which I call a *place selection.*

A place selection is defined by an infinite set of functions $s_n(x_1, x_2, \cdots x_{n-1})$ where $x_1, x_2, x_3, \cdots$ are the digits of an admissible number or a kollektiv and s_n has one of the two values zero or one. Here $s_n = 1$ means that the nth digit of the sequence is retained, $s_n = 0$ means that it is discarded. The decision about retaining or discarding the nth elements depends as you see, only on the preceding values $x_1, x_2, \cdots x_{n-1}$, but not on x_n or the following digits. Example of a place selection:

$$s_n = 1, \text{ if } x_{n-1} = 0 \text{ for prime numbers } n,$$

$$\text{if } x_{n-1} = 1 \text{ for } n \text{ not prime,}$$

$$s_1 = 1, \text{ and } s_n = 0 \text{ in all other cases.}$$

Experience shows that, if we apply such a place selection to the sequence X of outcomes of a game of chance, we get a new, selected sequence $S(X)$ in which the frequencies of gains and losses are about the same as in X. This fact or the practical *impossibility of a gambling system* suggests the adoption of the following procedure in handling transformations of admissible numbers.

First, if within a certain investigation the transformation applied to X is a place selection, we assume that the distribution in $X' = S(X)$ is the same as in X: distr $S(X)$ = distr X. Second, if a general transformation T is applied to X, say $X' = T(X)$, then we examine whether the existence of a place selection S that changes the distribution in X' (so as to have distr $S(X') \neq$ distr X') implies the existence of a place selection S_1 that would affect the distribution in X (so as to give distr $S_1(X)$ = distr X). If this is the case, we say that X' is a kollektiv, provided that the original sequence X was considered to be a kollektiv. Take e.g. for X the sequence resulting from tossing a die endlessly, and call $p_1, p_2, \cdots p_6$ the limiting frequencies of the six possible outcomes 1, 2, $\cdots$ 6 The transformation T may consist in replacing every 1 in the sequence X by a

2, every 3 by a 4, and every 5 by a 6. The new sequence consists of only three different kinds of elements 2, 4, 6 and therefore its distribution includes only three values p_2', p_4', p_6' where evidently $p_2' = p_1 + p_2$ etc. Here it is almost obvious that if a place selection applied to X' changes the value of p_2', the same selection if applied to X must change either p_1 or p_2. So, if the original sequence X was considered as a kollektiv, X' has to be admitted too.

Now the question arises whether this procedure is in itself consistent or whether it can lead to contradictions. We were concerned up to now with kollektivs the elements of which belong to a finite set of distinct numbers $e_1, e_2, \cdots e_k$ and the distributions of which are therefore defined by k non-negative values $p_1, p_2, \cdots p_k$ with the sum 1. In this case it was pointed out by Wald and by Copeland that, if an arbitrary distribution and an arbitrary countable set Σ of place selections are given, there exists a continuum of sequences every one of which has the given distribution, which is not affected by any place selection belonging to Σ. Now it may be supposed that in a concrete problem a sequence X' is derived from a sequence X by a finite number of fundamental operations involving a finite set Σ' of place selections. Another finite set Σ'' may consist of selections employed in establishing that certain sequences used in the derivation of X' are "combinable" ones. Finally an arbitrary countable set Σ of selections S may be assumed. According to our procedure we have shown that to any place selection S which affects the distribution in X' corresponds a certain S_1 which, when applied to X, changes the distribution of X. All these S_1 corresponding to the elements S of Σ form a countable set Σ_1. Now the set Σ_2 including Σ', Σ'', Σ_1 and also including all products of two of its own elements is a countable set too. What we use in computing the distribution of X' is only the fact that the given sequence X is unaffected by the selections that are elements of Σ_2. It follows from the above quoted results that we can substitute for X a numerically specified sequence and carry out all operations upon this specified sequence. So it is proved that no contradiction can arise in computing the final probability according to our conception.

I cannot enter here into a discussion of the more complicated case where the range within which the elements of a kollektiv vary, is an infinite one, either a countable set or a continuum. All principal problems connected with establishing the notion of kollektiv can be settled satisfactorily, at any rate, by considering those general forms of sequences as limiting cases of kollektivs with a finite set of attributes.

4. Example: Set-of-games problem. I want to present now a simple, but instructive example to show how the theory works and what task a mathematical foundation of the calculus of probability has to achieve. Let us recall the two sequences X and X' composed of zeros and ones of which we spoke above. The first represented the outcomes of a sequence of single games, the second the outcomes of triple sets of those games. If X is considered as a kollektiv with

given probabilities p and q for one and zero, it is easy to deduce the corresponding values p' and q' for X' and to show that X' is a kollektiv too We begin by carrying out three selections which single out from the original sequence x_1, x_2, x_3 $\cdots$ first, the elements x_1, x_4, x_7, $\cdots$ second, the elements x_2, x_5, x_8, $\cdots$ and third, the elements x_3, x_6, x_9, $\cdots$. It can be shown by means of certain further place selections that these three kollektivs which we call X_1, X_2, X_3 are combinable. That means that combining the corresponding elements of the three sequences like $x_1x_2x_3$, $x_4x_5x_6$, $x_7x_8x_9$, $\cdots$ leads to a new three dimensional kollektiv X_0 in which each permutation of three digits 0 and 1, has a probability equal to the corresponding product of p- and q-factors. For instance the probability of encountering the group 111 is p^3 and for the group 110 it is p^2q. Now we operate a mixing upon X_0 by collecting all permutations with two or three ones. We find in a well known way the sum $p^3 + 3p^2q$ for the probability p' of ones in the sequence X'. So far the result is very well known and can be reached—in my opinion, in a very incomplete and unsatisfactory way—also by the classical methods.

But what I want to discuss here is a slightly modified question. If the sequence X means gains and losses for single games and if the arrangement for sets of three games is made as indicated before, then in a real play the gains and losses of sets are counted in a different way. For, if the first two games of a set are both won or lost by the same player, the fate of the set is decided and there is no sense to play the third game. So the loss of the second set in our example will already be recognized after the fifth game and the actual sixth game will be considered as the first game of the third set. In this way the original sequence X decomposed into groups of two or three games

$$(X) \qquad 1\ 0\ 1\,|\,0\ 0\,|\,1\ 1\,|\,0\ 0\,|\,0\ 1\ 1\,|\,1\ 1\,|\,0\ 0\,|\,1\ 1\,|\,0\ 1\ 0\,|\,1\ 1\,|\,\cdots$$

leads to a new sequence X''

$$(X'') \qquad 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ \cdots$$

which is obviously different from X'. Everyone familiar with the usual handling of the probability concept will say that in X'' the probabilities of zeros and ones must be the same as in X' But a mathematical foundation of theory of probability, if it deserves this name, has to clear up the question: From what principles or particular assumptions and by what inferences may we deduce the equality of the limiting frequencies in X' and X''?

There is no difficulty in solving this problem from the point of view of the frequency theory. We have only to apply somewhat different place selections instead of the above used which lead to the kollektivs X_1, X_2, X_3. I showed elsewhere how the general set-of-games problem can be satisfactorily treated in this way Here I want to stress only that the problem as a whole is completely inaccessible by any of the other known approaches to probability theory. The classical point of view which starts with the notion of equally likely cases and rests upon a rather vague idea of the relationship between probability and

sequences of events does not even allow the formulation of the problem. In the so called modernized classical theory, as proposed by Fréchet, probabilities are defined as "physical magnitudes of which frequencies are measures." Fréchet would say that the frequencies both in X' and in X'' are measures of the same quantity. But why? We face here obviously a mathematical question which cannot be settled by referring to physical facts. It is clear that the equality of the distributions in the two sequences X' and X'' is due to the randomness or irregularity of the original sequence X. No theory which does not take in account the randomness, which avoids referring to this essential property of the sequences dealt with in probability problems, can contribute anything toward the solution of our question.

I have to make some special remarks about the so-called measure theory of probability.[2]

5. Probability as measure. Up to now we have been concerned only with the simplest type of kollektivs, namely, with those sequences the elements of which belong to a finite set of numbers so as to have a distribution consisting of a finite number of finite probabilities with the sum 1 It may be true that all practical problems, in a certain sense, fall into this range. For, the single result of an observation is always an integer, the number of smallest units accessible to the actual method of measuring. Nevertheless in many cases it is much more useful to adopt the point of view that the possible outcomes of an experiment belong to a more general set of numbers, e.g. to a continuous segment or any infinite variety. If we include the case of kollektivs of more than one dimension, we have to consider a point set in a k-dimensional space (where even k may be infinite) as the label set or attribute set of the kollektiv. In order to define the probability in this case we have to choose a subset A of the label set and to count among the first n elements the number n_A of those elements the attributes of which fall into A. Then the quotient $n_A \cdot n$ is the frequency, and its limiting value for n infinite will be called the probability of the attribute falling into A within the given kollektiv

It was rightly stressed by many authors that in the case of an infinite label set some additional restrictions must be introduced In particular A. Kolmogoroff set up a complete system of such restrictions. We cannot ask for the existence of the limiting frequency in any arbitrary subset A. It will be sufficient to assume that the limit exists for a certain Körper or a certain additive family of subsets. If it exists for two mutually exclusive subsets A and B, the limit corresponding to $A + B$ will be, by virtue of the original definition, the sum of the limits connected with A and B. We can now insert a further axiom involving the complete additivity of the limiting values. So we arrive at the statement

[2] What I call measure theory here is essentially that proposed by Kolmogoroff in his pamphlet of 1933. As to the new theory developed by Doob in his following paper (where instead of the label space the space of all logically possible sequences is used in establishing the measures) see my comment on page 215.

that probability is the measure of a set. All axioms of Kolmogoroff can be accepted within the framework of our theory as a part of it, but in no way as a substitute for the foregoing definition of probability.

Occasionally the expression probability as measure theory is used in a different sense. One tries to base the whole theory on the special notion of a set of measure zero. One of the basic assumptions in my theory is that in the sequence of results we obtain in tossing a so called correct die the frequency, say of the point 6, has a certain limiting value which equals 1/6 A different conception consists in stating that anything can happen in the long run with a correct die, even that an uninterrupted sequence of six's or an alternating sequence of two's and four's or so on may appear. Only all these events which do not lead to the limiting frequency 1/6 form, together as a whole, a set of events of measure zero. Instead of my assumption. the limiting value is 1/6 we should have to state: It is almost certain that a limit exists and equals 1/6 Nothing can be said against such an alluring assumption from an empirical standpoint, since actual experience extends in no case to an infinite range of observations. The only question is whether the asumption is compatible with a complete and consistent theory I cannot see how this may be achieved Before saying that a set has measure zero we have to introduce a measure system which can be done in innumerable ways. If e.g. we denote the outcome six by a one and all other outcomes 1 to 5 by zero, we get as the result of the game with a die an infinite sequence of zeros and ones. It has been shown by Borel that according to a common measure system the set of all 0, 1 sequences which do not have the limiting frequency $\frac{1}{2}$ has the measure zero. In this way it turns out to be almost certain that the limiting frequency of the outcome six in the case of a correct die is $\frac{1}{2}$. Other values for the limit can be obtained by a similar inference. It is a correct but misleading idea that the measure zero is unaffected by a regular (continuous) transformation of the assumed measure system, since in our field of problems different measures which are not obtained from one another by a regular transformation have equal rights So, saying that a certain set has the measure zero makes in our case no more sense than to state that an unknown length equals 3 without indicating the employed unit.

In recapitulating this paragraph I may say: First, the axioms of Kolmogoroff are concerned with the distribution function within one kollektiv and are *supplementary to my theory, not a substitute for it.* Second, using the notion of measure zero in an absolute way without reference to the arbitrarily assumed measure system, *leads to essential inconsistencies.*

6. Statistical estimation. Let me now turn to the last point, the application of probability theory to one of the most widely discussed questions in today's statistical research: the so-called estimation problem. Many strongly divergent opinions are facing each other here. I think that the probability theory based on the notion of kollektiv is best able to settle the dispute and to clear up the difficulties which arose in the controversies of different writers.

We may, without loss of generality, restrict ourselves to the simplest case of a single statistical variable x and a single parameter ϑ, where x of course may be the arithmetical mean of n observed values. Here (and likewise in the case of more variables and more parameters) we have to distinguish carefully among four different kollektivs which are simultaneously involved in the problem. The range within which both x and ϑ vary will be assumed to be a continuous interval so that all distributions will be given by probability densities.

The first kollektiv we deal with is a one-dimensional one where the probability of x falling into the interval x, $x + dx$ depends on x and on a parameter ϑ. If

$$p(x \mid \vartheta) \tag{1}$$

denotes the corresponding density and the limits A, B within which x possibly falls depend on ϑ too, we have

$$\int_{A(\vartheta)}^{B(\vartheta)} p(x \mid \vartheta)\, dx = 1 \qquad \text{for each } \vartheta. \tag{1'}$$

In order to fix the ideas we may imagine that the first kollektiv consists in drawing a number x out of an urn and that ϑ characterizes the contents of the urn. Asking for an estimate of ϑ implies the assumption that different possible urns are at our reach every one of which can be used for drawing the x. The ϑ values for the different urns fall into a certain interval C, D. It is usual to suppose that the urns are picked out at random so as to give another one-dimensional kollektiv with the independent variable ϑ. Let $p_0(\vartheta)\, d\vartheta$ be the probability of picking an urn with the characteristic value falling into the interval ϑ, $\vartheta + d\vartheta$. This density

$$p_0(\vartheta) \tag{2}$$

is often called the *prior* or *a priori* probability of ϑ. As the range within which ϑ varies is confined by the constants C and D, we have obviously

$$\int_C^D p_0(\vartheta)\, d\vartheta = 1. \tag{2'}$$

Now from these two one-dimensional kollektivs with the variables x in the first, ϑ in the second, we deduce by combination (multiplication) a two-dimensional kollektiv with the density function

$$P(\vartheta, x) = p_0(\vartheta) \cdot p(x \mid \vartheta). \tag{3}$$

The individual experiment which forms the element of this third kollektiv consists of picking at random an urn and drawing afterwards from this urn. Both x and ϑ are now independent variables (attributes of the kollektiv) and it is easy to see that it follows from (1) and (2)

$$\int_C^D \int_{A(\vartheta)}^{B(\vartheta)} P(\vartheta, x)\, dx\, d\vartheta = \int_C^D p_0(\vartheta)\, d\vartheta \int_{A(\vartheta)}^{B(\vartheta)} p(x \mid \vartheta)\, dx = 1. \tag{3'}$$

We will return later to this two-dimensional kollektiv. Let us, first, derive from it, by applying the operation of partitioning (Teilung), our fourth and last kollektiv which is one-dimensional again. Partitioning means that we drop from the sequence of experiments which form the third kollektiv all those for which the x-value falls outside a certain interval $x,\ x + dx$; and that in this way we consider a partial sequence of experiments with only the one variable ϑ. The distribution of ϑ-values within this sequence with quasi-constant x is given, according to the well known rule of division or rule of Bayes (a rule which can be proved mathematically) by[3]

$$p_1(\vartheta \mid x) = \frac{P(\vartheta, x)}{\int_C^D P(\vartheta, x)\, d\vartheta} = c(x)\, p_0(\vartheta)\, p(x \mid \vartheta). \tag{4}$$

It follows immediately that

$$\int_C^D p_1(\vartheta \mid x)\, d\vartheta = 1. \tag{4'}$$

This function p_1 of ϑ depending on the parameter x is generally called the *posterior* or *a posteriori* probability of ϑ.

If $p_1(\vartheta \mid x)$ can be computed according to the formula (4), every question concerning the "presumable" value of ϑ as drawn from the outcome x of an experiment is completely answered. We can find indeed, by integration the probability which corresponds to any part of the interval C, D of ϑ and so the estimation problem is definitely solved. But the trouble is that in most cases of practical application nothing or almost nothing is known about the prior probability $p_0(\vartheta)$ which appears as a factor in the expression of p_1. Hence arises the new question: *What can we say about the ϑ-values without having any information about its prior probability?* This is the estimation problem as it is generally conceived today.

The first successful approach to the answering of this question was made by Gauss. If we do not know p_1, we know however, except for a constant factor, the quotient p_1/p_0, posterior probability to prior probability which equals $cp(x \mid \vartheta)$. The maximum of this quotient must be greater than one, since the average values of both p_0 and p_1 are the same. So the maximum means the point of the greatest increase produced by the observed experimental value of x upon the probability of ϑ. It seems reasonable to assume the ϑ-value for which the ratio p_1/p_0 reaches its maximum as an estimate for ϑ: It is the value upon which the greatest emphasis is conferred by the observation. This idea, originally proposed by Gauss in his theory of errors, has been later developed chiefly by R. A. Fisher, and is known today as the maximum likelihood method. Calling the ratio p_1/p_0 likelihood seems indeed an adequate nomenclature.

[3] For brevity Bayes' rule is employed in the text as in the case of a discontinuous distribution. The correct procedure in the case of a continuous x would require that we first use finite intervals and then pass to the limit.

The method of estimation used most frequently today is not the maximum likelihood method, but the so called confidence interval method, inaugurated by R A. Fisher and now successfully extended and applied by J. Neyman. This method uses the third of the above mentioned kollektivs instead of the fourth, i.e. the two-dimensional probability $P(\vartheta, x)$. At first sight it seems hopeless to use this function which includes the unknown prior probability $p_0(\vartheta)$ as a factor. But it turns out as Neyman has shown[4] (and this is the decisive idea of the confidence interval method) that we can indicate in the x, ϑ-plane special regions for which the probability $\iint P(\vartheta, x)\, dx\, d\vartheta$ is independent of $p_0(\vartheta)$. In fact, if we point out for every ϑ such an interval x_1 , x_2 as to have

$$\int_{x_1(\vartheta)}^{x_2(\vartheta)} p(x \mid \vartheta)\, dx = \alpha, \qquad 0 < \alpha < 1, \tag{5}$$

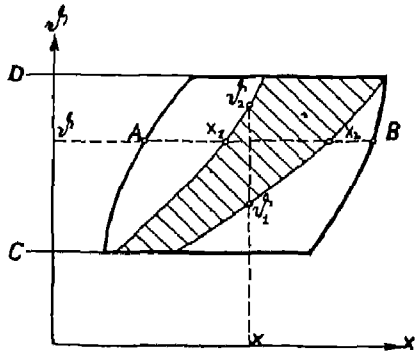


FIG. 1

it follows immediately from (2) and (5) for the region covered by these intervals

$$\int_C^D \int_{x_1(\vartheta)}^{x_2(\vartheta)} P(\vartheta, x)\, dx\, d\vartheta = \int_C^D p_0(\vartheta)\, d\vartheta \int_{x_1(\vartheta)}^{x_2(\vartheta)} p(x \mid \vartheta)\, dx = \alpha. \tag{6}$$

For given α the intervals can be chosen in different ways. If we choose $x_1 = A$ for $\vartheta = C$ and $x_2 = B$ for $\vartheta = D$, we get a strip or belt, as shown in Fig. 1 which supplies for every given x a smallest value ϑ_1 and a greatest value ϑ_2. The definition of our third kollektiv leads to the conclusion: *If we predict each time a certain x is observed that ϑ lies between the corresponding ϑ_1 and ϑ_2, then the probability is α that we are right, whatever the prior probability may be.*[5] It is

[4] J. Neyman, *Roy. Stat. Soc. Jour.*, Vol. 97 (1934), pp. 590-92

[5] After my lecture Dr A. Wald called my attention to Neyman's suggestion, namely that this statement can be generalized by admitting that the infinite sequence of ϑ-values which results from picking out successively the urns for drawing a number x, does not fulfill the conditions of a kollektiv. So, instead of the terms "whatever the prior probability may be" we can say "whatever the method of picking out the urns may be." In fact, let us consider the case where ϑ can assume only a finite number of values ϑ_1 , ϑ_2 , $\cdots$ ϑ_k . Among the n first trials let $n_\varkappa$ be the number of cases where $\vartheta = \vartheta_\varkappa$ and $n'_\varkappa \leqq n_\varkappa$ the number of cases where $\vartheta = \vartheta_\varkappa$ and x falls into the interval $x_1(\vartheta_\varkappa)$, $x_2(\vartheta_\varkappa)$. The relative

understood that in this argument both x and ϑ are variables the values of which may change from one trial to the next. I cannot agree with the statement, which is often made, that x only is a variable and ϑ a constant or that we are only interested in one specified value of ϑ. In no way is it possible, in the framework of the confidence limits method, to avoid the idea of a so-called superpopulation, i.e. the existence of a manifold of urns every one of which forms a kollektiv.[6] Thus no contradiction and no antagonism exists between this method and the Bayes formula. Only a different kollektiv, a two-dimensional instead of a one-dimensional, is here considered.

I have no time to enter here in a discussion of the very interesting developments of Neyman's theory which are intended to supply additional conditions in order to determine the arbitrary choice of the x-intervals in a unique way. May I only mention that what is called in Neyman's theory the probability of a second type error in testing the hypothesis $\vartheta = \vartheta_0$ is given by the expression

$$\int_C^D \int_{x_1(\vartheta_0)}^{x_2(\vartheta_0)} P(\vartheta, x)\, dx\, d\vartheta = \int_C^D p_0(\vartheta)\, d\vartheta \int_{x_1(\vartheta_0)}^{x_2(\vartheta_0)} p(x \mid \vartheta)\, dx. \tag{7}$$

If we want to determine the confidence belt or the intervals x_1, x_2 in such a way as to minimize this expression independently of the function $p_0(\vartheta)$, we obtain Neyman's maximum power condition

$$\int_{x_1(\vartheta_0)}^{x_2(\vartheta_0)} p(x \mid \vartheta)\, dx \equiv F(\vartheta, \vartheta_0) = \text{min. for each pair } \vartheta, \vartheta_0. \tag{8}$$

This condition, it is well known, cannot be fulfilled under general assumptions for $p(x \mid \vartheta)$. Moreover the above-mentioned boundary conditions $x_1(C) = A(C)$ and $x_2(D) = B(D)$ (or similar ones in other cases) have to be considered too. If they are not satisfied, the statement which can be made with probability α would include the prediction that certain x-values are impossible. Except for this case the above formulated theorem is equally valid for every region determined according to (5).

It is clear that if the original distribution is given by a regular, slightly varying function $p(x \mid \vartheta)$, the confidence limits method cannot give very substantial results. Let us take e.g. for $p(x \mid \vartheta)$ the uniform distribution

$$p(x \mid \vartheta) = 1/\vartheta \text{ for } 0 \leqq x \leqq \vartheta, \qquad 0 \leqq \vartheta \leqq 1. \tag{9}$$

---

frequency of correct predictions is then $(n_1' + n_2' + \cdots n_k'): n$ where n equals $n_1 + n_2 + \cdots n_k$. If n tends to infinity, at least one part of the n_κ must become infinite. For those the limit of $n_\kappa' : n_\kappa$ tends to α according (5) while the other terms (with finite n_κ and n_κ') have no influence. So the limiting value of the frequency $(n_1' + n_2' + \cdots n_k'): n$ equals in any event α. This generalization does not apply, if we ask for the probability of a second type error of the hypothesis $\vartheta = \vartheta_0$. Here the existence of the prior probability p_0 is essential.

[6] According to the generalization supplied by Neyman's point of view (*Phil. Trans. Roy. Soc.*, Vol. A-236 (1937), pp. 333–380) which is discussed in footnote 5, the superpopulation does not necessarily satisfy the conditions of a kollektiv.

We have here $A = 0$, $B = \vartheta$, $C = 0$, $D = 1$ and the domain in which x and ϑ vary is the 45° right triangle shown in Fig 2 Whatever $p_0(\vartheta)$ may be, the integral of $p(\vartheta, x) = p_0(\vartheta) \cdot p(x \mid \vartheta)$ over this domain is 1 and if we omit the part of the triangle on the left of the straight line $x = (1 - \alpha)\vartheta$, the integral over the remaining part is α. For $\alpha = 0\,90$, a statement which can be made with a probability of 90% reads The value of ϑ lies between x and $10x$ On the other hand we know from the very beginning with 100% certainty that ϑ lies between x and 1, so that for $x \geqq 0.1$ the statement is futile. (If one chooses as confidence belt the part on the left of the straight line $x = \alpha\vartheta$, the statement would run· ϑ lies between 1.1 x and 1 and values of x greater than 0 9 are impossible.) If we apply in this case the Bayes formula, we find that the outcome depends to the highest extent on what is known about the prior probability $p_0(\vartheta)$

In most cases however which present themselves in practical statistics the original density function $p(x \mid \vartheta)$ has a different character from that assumed in

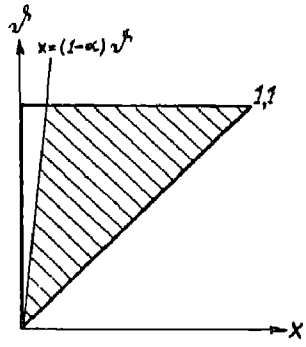


FIG. 2

(9). It depends generally on an integer n and the distribution is concentrated more and more when n increases. (We may define here concentration as standard deviation tending towards zero. The integer n means in general the number of basic experiments). We have e.g. in the so-called Bayes problem where x is the arithmetical mean of n observations the asymptotic expression for p:

$$p(x \mid \vartheta) \sim \sqrt{\frac{n}{2\pi\vartheta(1-\vartheta)}}\, e^{-\frac{1}{2}n(x-\vartheta)^2/\vartheta(1-\vartheta)} \tag{10}$$

$$0 \leqq \vartheta \leqq 1, \qquad 0 \leqq x \leqq 1.$$

If we denote by Φ the probability integral

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}\, du, \tag{11}$$

the x-intervals corresponding to a given probability value α are defined by

$$x_1 = \vartheta - \xi, \qquad x_2 = \vartheta + \xi \qquad \text{where } \Phi\left(\xi \sqrt{\frac{n}{2\vartheta(1-\vartheta)}}\right) = \alpha. \tag{12}$$

If n has a large value, the ξ's are very small and we get a narrow belt along the straight line $x = \vartheta$ as shown in Fig. 3 for $\alpha = 0.90$ and n about 100. The prediction which can be made with the probability α reads approximately

$$(13) \qquad x - \eta \leqq \vartheta \leqq x + \eta \qquad \text{where } \Phi\left(\eta \sqrt{\frac{n}{2x(1-x)}}\right) = \alpha.$$

On the other hand it is well known that in this case the Bayes formula supplies a posterior probability $p_1(\vartheta \mid x)$ which turns out to be more and more independent of the prior probability $p_0(\vartheta)$ when n increases. It has been shown that the asymptotic expression for $p_1(\vartheta \mid x)$ whatever $p_0(\vartheta)$ may be, is

$$(14) \qquad p_1(\vartheta \mid x) \sim \sqrt{\frac{n}{2\pi x(1-x)}}\, e^{-\frac{1}{2}n(\vartheta - x)^2/x(1-x)}.$$

It follows that, on the basis of the Bayes formula, we can predict for every single value of x with the probability α that ϑ lies between the above given

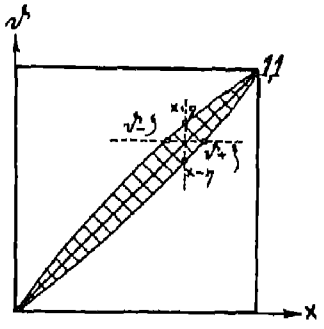


FIG. 3

limits (13). This is more than the confidence limits method supplies, but the result is subjected to the restriction that $p_0(\vartheta)$ is a continuous function. However, for large values of n (generally this means for large numbers of basic experiments) the outcomes of both methods are essentially the same.

Let me recapitulate in three brief sentences the essential results we have found in the problem of estimation.

1. There is no contradiction of any kind between the Bayes formula and the confidence limits method and no difference at all in the underlying probability concept. In both methods the idea of a sort of "super-population" is used. Only two different kollektivs are considered in both cases.

2. If the original distribution has a regular, slightly varying density function $p(x \mid \vartheta)$, the Bayes method gives a complete answer when the prior probability is known and no answer when it is unknown. The confidence limits method gives in both cases a definite solution; it lies in the nature of things that the solution cannot be very substantial if $p(x, \vartheta)$ is only slightly varying.

3. If the original distribution $p(x \mid \vartheta)$ depends on a further parameter n and becomes concentrated more and more with increasing n, both approaches give, for large n, asymptotically about the same results.

It is not intended by these remarks to impair the value of the confidence limits method which both from theoretical and from practical point of view deserves our attention. But the rather inconceivably aggressive attitude towards the Bayes' theory as displayed by a number of statisticians, which, however, does not include J. Neyman, turns out to be completely unfounded.

PROBABILITY AS MEASURE

By J. L. Doob

University of Illinois

The following pages outline a treatment of probability suitable for statisticians and for mathematicians working in that field. No attempt will be made to develop a theory of probability which does not use numbers for probabilities. The theory will be developed in such a way that the classical proofs of probability theorems will need no change, although the reasoning used may have a sounder mathematical basis It will be seen that this mathematical basis is highly technical, but that, as applied to simple problems, it becomes the set-up used by every statistician. The formal and empirical aspects of probability will be kept carefully separate. In this way, we hope to avoid the airy flights of fancy which distinguish many probability discussions and which are irrelevant to the problems actually encountered by either mathematician or statistician.

We shall identify as Problem I the problem of setting up a formal calculus to deal with (probability) numbers. Within this discipline, once set up, the only problems will be mathematical. The concepts involved will be ordinary mathematical ones, constantly used in other fields. The words "probability," "independent," etc will be given mathematical meanings, where they are used.

We shall identify as Problem II the problem of finding a translation of the results of the formal calculus which makes them relevant to empirical practice. Using this translation, experiments may suggest new mathematical theorems. If so, the theorems must be stated in mathematical language, and their validity will be independent of the experiments which suggested them. (Of course, if a theorem, after translation into practical language, contradicts experience, the contradiction will mean that the probability calculus, or the translation, is inappropriate.)

The classical probability investigators did not separate Problems I and II carefully, thinking of probability numbers as numbers corresponding to events or to hypothetical truths, and always referring the numbers back to their physical counterparts. The measure approach to the probability calculus has put this approach into abstract form, and separated out the empirical elements, thus removing all aspects of Problem II. We shall explain this approach first in a simplified set-up, that which will be made to correspond (Problem II) to a repeated experiment in which the results of the nth trial can be any integer x_n between 1 and N (inclusive), in which the experiments are independent of each other, and performed under the same conditions. (The set-up will be applicable, for example, to the repeated throwing of a die.)

The measure approach treats this experiment as follows. Let $\omega: (x_1, x_2, \cdots)$ be any sequence of integers between 1 and N, inclusive We consider ω as a point in an infinite dimensional space Ω (Each point ω may be considered as a logically possible sequence of results of the given experiment, and this fact will guide us in solving Problem II) A measure function is defined on certain sets of points of Ω as follows. Let $p_1, \cdots, p_N$ be any numbers satisfying the conditions

$$p_j \geqq 0, \qquad j \geqq 1, \qquad p_1 + \cdots + p_N = 1.$$

(How these numbers are chosen in any particular problem will be explained below. The method of choice is irrelevant to the mathematics, but is involved in the solution of Problem II.) The set of all sequences beginning with $x_1 = \alpha$ is given measure p_α. More generally, the measure of the set of all sequences beginning with $x_1 = \alpha_1, \cdots, x_n = \alpha_n$, is defined as $p_{\alpha_1} \cdot p_{\alpha_2} \cdots p_{\alpha_n}$. In this way, as can be shown,[1] a completely additive measure function is determined on certain point sets of Ω, on a field $\mathfrak{F}$ of sets so large that all the usual Lebesgue measure and integration theory is applicable. This means that there is a collection $\mathfrak{F}$ of sets of points of Ω such that if $S_1, S_2, \cdots$ are finitely or infinitely many sets in the collection, their sum $\sum_1^\infty S_n$, their intersection $\prod_1^\infty S_n$, and their complements are also in the collection Each set S in $\mathfrak{F}$ has a definite measure $P(S)$, $0 \leqq P(S) \leqq 1$, and if $S_1, S_2, \cdots$ are finitely or infinitely many disjunct sets in $\mathfrak{F}$,

$$P(S_1 + S_2 + \cdots) = P(S_1) + P(S_2) + \cdots.$$

Problem II, the translation problem, is solved as follows. Each relevant event is made to correspond to a point set of Ω. A relevant event is a physical concept—defined by imposing some set C of conditions on the results of the experiments. The corresponding Ω-set is the set of sequences $(x_1, x_2, \cdots)$ satisfying the same set C of conditions, imposed on the x_j. Thus the set of all sequences beginning with $x_1 = \alpha_1, x_2 = \alpha_2$, is made to correspond to the event. *the result of the first experiment is* α_1, *of the second is* α_2. As is to be expected, the mathematical picture goes further than the real one The "event" *1 occurs infinitely often in a sequence of trials* has only conceptual significance, physically, but the corresponding point set of Ω. the set of all sequences $(x_1, x_2, \cdots)$ containing infinitely many 1's, is a perfectly definite point set whose measure can be calculated in terms of $p_1, \cdots, p_N$. (In fact it is easily seen that this measure is 1 or 0, according as $p_1 > 0$ or $p_1 = 0$.) By "the probability of an event" we shall mean the measure of the corresponding Ω-set. As this measure has been defined, the probability that the nth trial results in a number j is p_j, and the probability that one trial results in j, and another in k, is $p_j \cdot p_k$.

[1] Cf A. Kolmogoroff, *Ergebnisse der Mathematik*, Vol. 2, No. 3, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, where the most complete treatment of the approach to the probability calculus from the standpoint of measure is given.

The justification of the above correspondence between events and Ω-sets is that certain mathematical theorems can be proved, filling out a picture on the mathematical side which seems to be an approximation to reality, or rather an abstraction of reality, close enough to the real picture to be helpful in prescribing practical rules of statistical procedure. The following two theorems are important ones, from this point of view. These two theorems depend in no way on observed facts. They are stated and proved in the customary language of modern analysis.

THEOREM A. Let j_n be the number of the first n coordinates of the point $\omega:(x_1, x_2, \cdots)$ which are equal to j, where j is some integer $(1 \leqq j \leqq N)$ which will be kept fixed throughout the discussion. Then $0 \leqq j_n \leqq n$, and j_n varies from point to point on $\Omega: j_n = j_n(\omega)$ is a function of ω, that is of the sequence $(x_1, x_2, \cdots)$. When $n \to \infty$, j_n/n has not a unique limit independent of the sequence $(x_1, x_2, \cdots)$ under consideration. In fact if ω is the point $(k, k, \cdots)$, $j_n(\omega) = 0$ for all n, unless $j = k$; if ω is the point $(j, j, \cdots)$, $j_n(\omega) = n$ for all n. It is simple to give examples of sequences $\omega:(x_1, x_2, \cdots)$ for which $j_n(\omega)$ oscillates without approaching a limit, as $n \to \infty$. But Theorem A (usually called the strong law of large numbers) states that there is a set of sequences, i.e. an ω-set S, of measure 0, such that

$$\lim_{n\to\infty} \frac{j_n(\omega)}{n} = p_j, \tag{1}$$

unless ω is in S. In other words the sequences for which (1) is not true are exceptional in the sense of measure theory. If a new choice $\{p_i'\}$ of p_i's is made, then if $p_j' \neq p_j$, the new exceptional set includes all the sequences which were not exceptional before, since the limit in (1) becomes p_j'. Thus S depends essentially on p_j. Theorem A is a generalization of Bernoulli's classical theorem which states in our language that the measure of the set of sequences $\omega:(x_1, x_2, \cdots)$ for which

$$|j_n(\omega)/n - p_j| > \epsilon$$

approaches 0, as $n \to \infty$, for any positive ϵ. Theorem A is stronger because it states that there is actual convergence, whereas Bernoulli's theorem only concludes that there is a kind of convergence on the average.

Theorem A corresponds to certain observed facts, relating to the clustering of "success ratios," giving rise to empirical numbers $\bar{p}_j$. If the statistician wishes to apply his calculus to a given experiment (Problem II), he sets $p_i = \bar{p}_j$. There has been frequent discussion of the problem of determining the $\bar{p}_j$. This discussion of the $\bar{p}_j$ is sometimes held on so high a plane that the innocent bystander may wonder to what purpose such abstract philosophic concepts could possibly be put—besides that of stimulating further discussion on a still higher plane. The principle purpose of this paper is to discuss Problem I, but a few words on Problem II might not be out of place here. Almost everyone who is going to use probability numbers, the $\bar{p}_j$, *for other than conversational purposes,*

derives them in the same way. There is a judicious mixture of experiments with reason founded on theory and experience. Thus if a coin is tossed by an experimenter who has examined the coin, and found that it had heads on one side but not on both, that it seemed balanced, and that (as a confirming check) tossing a hundred times gave around 50 heads, the experimenter would use $\frac{1}{2}$ as the probability of obtaining heads in his further reasoning. Of course there is no logic compelling this. The experimenter may have been fooled. A coin far out of balance may turn up 50 heads in 100 throws. But man must act, and the above procedure has been found useful, which is all that is desired In many experiments, less reliance can be placed on a preliminary physical examination of the experimental conditions, and more must be placed on the actual working out of the experiment, as in the analysis of machine products. In that case, the actual results must be examined with great care, before attempting to use the above mathematical set-up. It sometimes may even be possible to change the experimental conditions to make the mathematics applicable.[2] In all cases, such mathematical theorems as Theorem A and the following Theorem B give the basis for applying the formal apparatus to practice. Indeed, the criterion of application includes the verification of special cases of the practical versions of Theorems A and B

THEOREM B: Let $f_n(x_1, \cdots, x_{n-1})$ $(n > 1)$ be any function of the indicated variables, except that we suppose f_n only takes on the values 0, 1 Let $\omega: (x_1, x_2, \cdots)$ be a given point of Ω Let n' be the number of the first n integers i such that $f_i(x_1, \cdots, x_{i-1}) = 1$, and let j_n' be the number of the first n integers i such that $f_i(x_1, \cdots, x_{i-1}) = 1$, and $x_i = j$. Then j_n', n' are functions of $\omega: (x_1, x_2, \cdots)$. If $f_1 \equiv f_2 \equiv \cdots \equiv 1$, $j_n' = j_n$, $n' = n$, where j_n is as defined above. Suppose that there is an Ω-set S_0 of measure 0 such that $n' \to \infty$, as $n \to \infty$, unless $\omega \in S$ Theorem B states that there is then an Ω-set S' of measure 0, such that if $\omega: (x_1, x_2, \cdots)$ is not in S',

$$\lim_{n\to\infty} \frac{j_n'(\omega)}{n'} = p_j. \tag{1'}$$

(The set S' will depend on the given functions $f_1, f_2, \cdots$ and on the p_i, but is fixed, once these have been chosen.) This mathematical theorem corresponds to certain observed facts (usually summarized by stating that no (successful) system of play is possible). In fact, it states, in the language of practice, that rejecting certain trials, using as a criterion of acceptance or rejection the results of preceding trials, rejecting the ith trial if $f_i(x_1, \cdots, x_{i-1}) = 0$, does not affect the outcome of a game of chance, or, more precisely, does not affect the validity of the physical fact corresponding to Theorem A. If $f_1 \equiv f_2 \equiv \cdots \equiv 1$, (1') becomes (1). The hypothesis that $n' \to \infty$ as $n \to \infty$ unless $\omega \in S_0$ is made to insure that infinitely many trials will be accepted. As an example of the

[2] Cf. W. A. Shewhart, *Statistical Method from the Viewpoint of Quality Control*, Washington, 1939.

possible variety in the definition of the f_i, we might define f_i as 1 if $x_{i-1} = N$, and $f_i = 0$ otherwise, so trials are accepted only if the previous trial resulted in the number N. Or much more complicated systems can easily be devised in which the criterion of acceptance of the nth trial depends on a varying number of the results of preceding trials. This theorem gives a mathematical counterpart to the physical idea of the mutual independence of repeated trials

To summarize, mathematically (Problem I) the study has been reduced to that of the measure properties of Ω. This can be considered independently of any physical correspondence. The physical correspondence (Problem II) makes any event $\mathfrak{E}$ correspond to a point set E of Ω, the "probability of $\mathfrak{E}$" becomes the measure of E. Thus "the probability that the result of the first experiment is 3" becomes the measure of the set of sequences $(x_1, x_2, \cdots)$ beginning with $x_1 = 3$. *We have given no sharp definition of probability as a physical concept.* If the above mathematical set-up, after translation, using some set of p_j's, seems to fit a given physical set-up, any event will be said to have as its probability, the measure of the corresponding Ω-set. We have attempted to give no intrinsic a priori definition of the probability of an event: such a definition is quite unnecessary for our purposes. All that was required was a basis for prescribing the usual statistical procedures, and we have described such a basis.

In the above example, there would have been no new difficulty introduced if the x_n were not restricted to integral values, but allowed to take on any numerical values. The general point $\omega: (x_1, x_2, \cdots)$ of Ω would now be any sequence of real numbers. Instead of choosing the numbers $p_1, \cdots, p_N$ we choose a "distribution function" $F(x)$, a monotone function with the following properties.

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1, \qquad F(x - 0) = F(x).$$

Measure on Ω is defined as follows. The set of all sequences beginning with x_1 such that $a \leqq x_1 < b$ is given measure $F(b) - F(a)$. (The number $F(b)$ is called "the probability that $x_1 < b$.") More generally, the measure of the set of all sequences $(x_1, x_2, \cdots)$ beginning with $x_1, \cdots, x_n$, such that $a_j \leqq x_j < b_j$, $j = 1, \cdots, n$ is defined as $\prod_j [F(b_j) - F(a_j)]$. Thus if $F(x)$ defines a simple rectangular distribution: $F(x) = 0$ for $x < 0$, $F(x) = x$ for $0 \leqq x \leqq 1$, $F(x) = 1$ for $x > 1$, Ω-measure becomes (infinite dimensional) volume in the (infinite dimensional) unit cube. The correspondence (Problem II) between events and point sets of Ω is defined just as before. Sometimes it may be useful, in considering experiments giving rise to pairs of numbers, to let each x_n be a pair of numbers so that Ω becomes a sequence of points of a plane instead of a sequence of points of a line. In all cases there are mathematical theorems true of the resulting Ω which guide us (Problem II) in deciding just how the Ω-measure is to be defined, that is, how $F(x)$ is to be defined, in dealing with a given practical problem. But the essential point is this. Once Ω-measure has been defined, no changes or further hypotheses are possible or necessary. All

relevant probability questions are answerable. Thus consider a question of the following type if the experiments are grouped in some way,[3] with what probability will the groups have some given regularity property?[4] The question singles out a set E of sequences of Ω and asks: what is the measure of E? The problem may or may not be difficult mathematically,[5] depending on the grouping, but the original definition of measure on Ω needs no enlargement to answer it.

Technically, the mathematics has become the mathematics of a special type of measure defined on a space of infinitely many dimensions. If, however there is an integer ν such that only at most ν experiments are to be considered, we need only consider the ν-dimensional space of points $(x_1, \cdots, x_\nu)$, defining measure in this space in the same way as on Ω. Thus if x_n has the rectangular distribution defined above, the measure in $(x_1, \cdots, x_\nu)$-space becomes ordinary ν-dimensional volume in the unit cube. Perhaps the most common measure a statistician considers is that in which the measure of an $(x_1, \cdots, x_\nu)$-set E becomes "the probability that the point $(x_1, \cdots, x_\nu)$ representing an independent sample of ν from a normal distribution of mean 0 and variance σ^2" will lie in E.

$$P\{E\} = \sigma^{-\nu}(2\pi)^{-\frac{1}{2}\nu} \int \cdots \int_E e^{-\frac{1}{2}(x_1^2 + \cdots + x_\nu^2)/\sigma^2} dx_1 \cdots dx_\nu . \tag{2}$$

This example makes it obvious that the statistician is always doing measure theory, even though he may not state that fact explicitly. If the number of experiments has no upper bound conceptually—mathematically when the number of dimensions ν may increase without limit, as in Theorems A, B, it is much more convenient to use the space Ω, in terms of which experiments with varying numbers of trials can be considered simultaneously. The classical proofs of probability theorems, such as Bernoulli's theorem (the law of large numbers) are perfectly correct If the "probability of an event" is interpreted as the measure of a set, these proofs do not even need verbal changes. There can be no question of the need for any axiomatic development beyond that necessary for measure theory, and the probability calculus can lead to no contradiction, unless the theory of measure is faulty.

It is customary for probability theorists to stop their discussions when the present stage is reached, so that the beginnings of a formal calculus have been constructed to deal with a repetition of independent experiments, conducted

[3] A grouping is necessary, for example, when two players are playing a game in which two out of three wins in the trials win a game. The trials are then grouped into successive groups of two or three, depending on how they come out

[4] Continuing the preceding note, the question might be: will the ratio (games won by player α)/(games played) approach a limit with probability 1, that is, for all of the original sequences $\{x_n\}$ except possibly some forming a set of measure 0?

[5] The answer to the question of the preceding notes is simple. If p is the probability that player α wins a trial, the ratio in question approaches $p^3 + 3p^2(1 - p)$, the probability that α wins a game, with probability 1

under the same conditions. Perhaps this is because of the following widely held syllogism: probability is something dealing with random events; random events are events having no influence on each other; therefore .... Unfortunately mathematicians and statisticians must deal with many problems involving dependent probabilities, whose solutions require the most delicate and careful applications of modern analysis. The rudimentary calculi which the outsiders find esthetically or philosophically pleasing are usually either insufferably awkward or completely insufficient for the needs of professionals. There is a strange situation, which one observer has facetiously described somewhat as follows: it is true with probability 1 that the technical workers in probability use the measure approach, but that the writers on "probability in general" descendants of Carlyle's professor, do not consider this approach worth much more than a passing remark.[6] The following pages outline how our previous treatment is generalized to deal with problems in which it is desirable to have the distribution of x_j vary with j (so that physically the experiments are no longer the same), and in which the x_j do not have to correspond to the results of independent experiments. Some attempt will also be made to show how the modern mathematical theory of real functions is applied to the probability calculus.

Let $x_j = x_j(\omega)$ be the jth coordinate of the point $\omega: (x_1, x_2, \cdots)$. Then as the sequence $\omega: (x_1, x_2, \cdots)$ varies, x_j does also: $x_j(\omega)$ is a function of ω. The functions $x_1(\omega)$, $x_2(\omega)$, $\cdots$ are functions defined on Ω, an abstract space on which a measure has been defined. Moreover Ω-measure has been defined in such a way that the Ω-set for which $x_j(\omega) < K$ (j, K fixed) is an Ω-set whose measure has been defined (This set is composed of all sequences $(x_1, x_2, \cdots)$ whose jth coordinate is $<K$, and the measure is $F(K)$, using our last definition of Ω-measure.) In the terminology of measure theory, $x_j(\omega)$ is thus a measurable function. The study of the measure relations of Ω, and this is the whole of our probability calculus, can be considered, from this point of view, as the study of the properties of a sequence of measurable functions, one with very special properties, as we shall see, defined on some space. A measurable function defined on Ω is usually called a chance variable, in the theory of probability. (This terminology is somewhat dangerous, because it mixes Problems I and II.) The whole apparatus of modern real variable theory is applicable to these chance variables. Thus if $f(\omega)$ is a chance variable (measurable function of ω) (physically, a function of the observations), it is customary to define a number called its expectation. This number is simply the integral of $f(\omega)$, with respect to the given Ω-measure. The fact that the expectation of the sum of two chance variables is the sum of their expectations is simply the familiar theorem that the integral of the sum of two functions is the sum of their integrals. Let $S(j, K)$ be the Ω-set defined by the inequality $x_j < K$. Up to now we have supposed

[6] This analysis, like every other probability statement, is only an approximation to reality, but a fairly close one.

that the measure of $S(j, K)$ is independent of j, that is that the distribution of x_j is independent of j. We have also supposed that[7]

$$P\{S(1, K_1) \cdots S(n, K_n)\} = P\{S(1, K_1)\} \cdots P\{S(n, K_n)\} \tag{3}$$

for any positive integer n, and numbers $K_1, \cdots, K_n$. That is, we have supposed that $x_1(\omega)$, $x_2(\omega)$, $\cdots$ are mutually independent chance variables.[8] In fact probability measure on Ω has been defined just to make the foregoing two facts true. Mutual independence is a very strong hypothesis to impose on a sequence of functions In many probability problems (Markoff chains for example), more general measures must be defined on Ω. The sequence $x_1(\omega)$, $x_2(\omega)$, $\cdots$ whose properties are those of Ω-measure, is then no longer a sequence of independent functions, and the distribution of x_j can vary with j.

At this level, the study becomes the study of any sequence of measurable functions, defined on some space of total measure 1. If f, g are given chance variables, they may turn out to be independent. In that case the theorem that the expectation of their product is the product of their expectations becomes, when translated into mathematical language, the familiar theorem that

$$\int\int f(x)g(y)\,dx\,dy = \int f(x)\,dx \int g(y)\,dy.$$

The mathematical theorems are not simply analogues of the probability theorems—they themselves are those theorems. When stated mathematically, the probability theorems need no proof: they need only recognition as standard results.

Empirical needs suggest that certain functions called conditional probability distributions, and conditional expectations, should be defined in a certain way. This is possible, as a formal matter,[9] and the theorems then proved about these functions gives them their usual meaning when translated into practical language. These functions are extremely useful tools in dealing with mutually dependent (that is not independent) chance variables

The above approach is easily generalized to the stage needed in the study of Brownian movements or of time series, in which, instead of the proper initial

---

[7] $P\{S\}$ was defined as the measure of the Ω-set S.

[8] The n chance variables $f_1(\omega), f_2(\omega), \cdots, f_n(\omega)$ are said to be independent if for every set of n numbers $K_1, \cdots, K_n$, the following equality is true.

$$P\{f_j(\omega) < K_j, \quad j = 1, \cdots, n\} = \prod_j P\{f_j(\omega) < K_j\},$$

where $P\{\cdots\}$ denotes the Ω-measure of the Ω-set defined by the conditions in the braces. Thus in the example of a normal distribution in ν dimensions given above, $x_1, \cdots, x_\nu$ are independent functions on the space of ν dimensions, a fact which follows readily from the fact that the ν-dimensional density function is the product of ν functions of the separate variables.

[9] Cf. Kolmogoroff, loc. cit

abstraction being a sequence $\{x_n\}$ of numbers, we have a one-parameter family $\{x_t\}$ (t takes on all real values). The number x_t may, for example, be thought of as the x-coordinate of a particle at time t. There is no difference in principle here: Ω is now the space of functions of t, instead of the space of sequences, that is functions of n. From the other point of view, instead of studying the properties of a sequence of measurable functions, it becomes necessary to study the properties of a one-parameter family of measurable functions.

DISCUSSION OF PAPERS ON PROBABILITY THEORY

By R. von Mises and J. L. Doob

1. **Comments by R. von Mises.** Professor Doob outlines a new theory of probability starting with the following three basic conceptions. First, he uses the notion of an infinite sequence of trials or better, of an infinite sequence of numbers $x_1, x_2, x_3, \cdots$ which can be considered as the outcomes of infinitely repeated uniform experiments. Second, he introduces (in his Theorem A) the limit of the relative frequency of a particular outcome α. Third, (in his Theorem B) the notion of place selection defined by a sequence of functions $f_n(x_1, x_2, \cdots x_{n-1})$ is employed. All these three concepts are completely strange to the so called classical theory as developed by Bernoulli, Laplace, Poisson, etc. They have been introduced and made the corner stone of probability theory in my papers published since 1919. I daresay that in no probability investigation before 1919 any of those notions even were mentioned.

This concerns what Professor Doob calls the Problem I or the purely mathematical aspect of the question. As to his Problem II or the relationship between the formal calculus and real facts Professor Doob stresses that the actual values for probabilities that enter as data into a particular argument have to be drawn from long, finite sequences of experiments. This is in complete accordance with the standpoint of my theory and in strict contradiction to the classical conception which knows only "a priori" probabilities determined by "equally likely cases."

In both theories, Professor Doob's and mine (not in the classical) a mathematical model or picture is associated with a long sequence of uniform experiments. These models are different in both theories. My model (the "kollektiv") consists of one infinite sequence ω: $x_1, x_2, x_3, \cdots$ in which the limit of the relative frequency of each possible outcome α exists and is indifferent to a place selection; the value of this limit is called the probability of α.

On the other hand Professor Doob's model implies all logically possible sequences which form a space Ω and he shows that in this space a measure function can be introduced which fulfills the following conditions: (1) If m is a positive integer, the set of all sequences the mth element of which is α has a measure p_α independent of m; (2) the set of all sequences in which the relative frequency of α-results has either no limit or a limit different from p_α is zero; (3) if S is any place selection, the set of all sequences ω for which the relative frequency of α in $S(\omega)$ has either no limit or a limit different from p_α is likewise zero; this value p_α is called the probability of the outcome α. It then can be shown that a probability in this sense can be ascribed to certain events, i.e. to certain types of experiments which in some way are connected with the sequence of basic

experiments. E.g. if the original sequence consists of the single successive tossings of a die, the derived sequence may consist of pairs of tossings with the sum of the outcoming points as new value of α. The new probabilities p'_α are found as measures of certain sets in the original measure system established in Ω.

There is no doubt that the model used by Professor Doob for representing empirical sequences of uniform experiments is logically consistent. Its practical usefulness depends on how the usual problems of combining different kollektivs and so on can be settled within this scheme. This has to be shown in detail. It seems to me that my conception is simpler in its application and closer to reality, while his model may be considered more satisfactory from a logical standpoint since it avoids the difficulties connected with the concept of "all place selections." At any rate, however, there is no contradiction or irreconcilable contrast: both theories are essentially statistical or frequency theories, equally far from the classical conception based on "equally likely cases." In both theories probabilities are, of course, measures of sets.

2. **Comments by J. L. Doob.** It is perhaps unfortunate that Professor von Mises' treatment of probability problems, based on typical sequences ("collectives," "admissible numbers"), is commonly called the "frequency theory."[1] It is clear to any reader of our papers (identified as M and D below) that the idea of frequency, at least in the discussion of the relation of mathematics to practice, is no more fundamental to one approach than to the other. In one mathematical treatment frequency notions first appear in the theorems, whereas in the other they first appear in the axioms; but they appear in both. The principal objection the measure advocates have to the frequency approach is that it is awkward mathematically. Anyone who doubts this awkwardness need only examine various books published recently, using this approach, to see what a lot of fussy detail is involved merely in proving such elementary results as the Tchebycheff inequality or the Bernoulli theorem. One author considers it necessary to have his chance variables so restricted that if x is a chance variable, the event $x < k$ has a probability assigned to it only if k is not in some exceptional set, which may be infinite. To take another example, consider the coin tossing game discussed in both M and D, in which two out of three wins at tosses win a game. Apparently the probability analysis of this game is somewhat difficult in terms of the frequency theory. As the quite elementary treatment outlined in D shows, there is no difficulty involved, using the measure approach. The question is simple: a set of chance variables is given (corresponding to the original tosses); a new set is determined from them (corresponding to the grouping into games). Only elementary algebraic manipulation is required to verify that the new chance variables are mutually independent in the mathematical sense, (Cf. D), and have the same distribution, so the law of large numbers is applicable. Professor von Mises considers that the measure theory cannot handle this problem. I on the other hand consider that this problem exhibits the mathematical disadvantages of the frequency theory.

[1] This identifying name will be used below also.

The frequency theory reduces everything to the study of sequences of mutually independent chance variables, having a common distribution. "Probability theory is the study of the transformations of admissible numbers" writes Professor von Mises. This point of view is extremely narrow Many problems of probability, say those involved in time series, can only be reduced in a most artificial way to the study of a sequence of mutually independent chance variables, and the actual study is not helped by this reduction, which is merely a tour de force.

It is claimed in M that the axioms of measure theory only describe the distribution within one collective (M, p. 00). This statement seems to mean that only the measure relations (using the notation of D) of the first coordinate function $x_1(\omega)$ can be discussed in the measure theory, that is only probabilities of the type: the probability that $x_1 < k$ (in the language of practice, "the probability that the result of the first experiment is less than k") are discussed. Actually, however, (Cf. D) the measure theory can discuss any number of experiments simultaneously, using the appropriate space Ω

Many of the debates between the advocates of the various probability theories have been wasted, because some of the debaters talk mathematics, others physics. With this in mind, I should like to stress again[2] that (except for a few philosophically inclined Englishmen) everyone calculates probability numbers in the same way—a combination of reasoning based on experience and helped by theory, with examination of the experimental conditions and the results of trials. Frequency considerations necessarily play a large part The fact that almost everyone calculates probability numbers in the same way does not alter the fact that one mathematical theory may be more useful or convenient than another in dealing with these probability numbers.

In closing, it seems proper to call attention to what the measure advocates consider the real services and contributions of the approach of Professor von Mises. Professor von Mises was the first to stress the importance of the second of two fundamental generalizations of experience in dealing with repeated mutually independent experiments of the same character: (1) the clustering of success ratios and (2) the fact that this clustering is unaffected by a system of rejection as described in M and D. These two generalizations of experience are certainly fundamental. The only point under discussion here is how such generalizations are to be put into a mathematical setting. The original such setting of Professor von Mises was criticized as not really mathematical. The setting now proposed by Copeland and others is criticized by the measure advocates as mathematically inflexible and clumsy. But it is significant that even in a treatment of the measure approach, as in D, it was felt essential to stress the mathematical interpretation of the two empirical generalizations of Professor von Mises In the terminology of D, the measure advocates consider the contribution of Professor von Mises' approach to be a contribution to a solution of Problem II, not to Problem I, the mathematical problem.

---

[2] We are not talking mathematics now, but the application of mathematics

CONTINUED FRACTIONS FOR THE INCOMPLETE BETA FUNCTION[1]

By Leo A. Aroian

Hunter College

1. Introduction. Existing literature on the problem of calculating the incomplete Beta function

$$B_x(p, q) = \int_0^x x^{p-1}(1-x)^{q-1}\,dx, \qquad 0 < x < 1, p > 0, q > 0, \tag{1.1}$$

and the levels of significance of Fisher's z [1] leave further work to be done. Muller's continued fraction and a new continued fraction are shown to possess complementary features covering the range of $B_x(p, q)$ for all values of x, p, q. Previous methods of computing $I_x(p, q) = B_x(p, q)/B(p, q)$ are given in [2], [5], [6], [8], [10], [13], [14], [15].

Müller's continued fraction is

$$I_x(p, q) = C\left[\frac{b_1}{1+}\,\frac{b_2}{1+}\,\frac{b_3}{1+}\,\frac{b_4}{1+}\,\cdots\right], \tag{1.2}$$

where

$$C = \frac{\Gamma(p+q)}{\Gamma(p+1)\Gamma(q)}\,x^p(1-x)^{q-1}, \qquad b_1 = 1, \qquad \mu_s = \frac{q-s}{p+s},$$

$$b_{2s} = -\frac{(p+s-1)(p+s)}{(p+2s-2)(p+2s-1)}\,\mu_s\,\frac{x}{1-x},$$

$$b_{2s+1} = \frac{s(p+q+s)}{(p+2s-1)(p+2s)}\,\frac{x}{1-x}.$$

A convergent infinite series $1 + \sum_{n=1}^{\infty} d_n x^n$ can be converted into an infinite continued fraction of the form $\frac{1}{1+}\,\frac{c_1 x}{1+}\,\frac{c_2 x}{1+}\,\cdots$ where [4], [9] p. 304,

$$c_1 = -\beta_1, \qquad c_2 = \frac{-\beta_2}{\beta_1},$$

$$c_{2s} = \frac{-\beta_{2s-3}\beta_{2s}}{\beta_{2s-2}\beta_{2s-1}}, \qquad c_{2s+1} = \frac{-\beta_{2s-2}\beta_{2s+1}}{\beta_{2s-1}\beta_{2s}}, \qquad s > 2 \tag{1.3}$$

[1] Presented at a meeting of the American Mathematical Society, October 28, 1939, New York City.

where

$$\beta_{2s} = \begin{vmatrix} 1 & d_1 & d_2 & \cdots & d_s \\ d_1 & d_2 & d_3 & \cdots & d_{s+1} \\ \vdots & & & & \vdots \\ d_s & d_{s+1} & d_{s+2} & \cdots & d_{2s} \end{vmatrix}, \qquad \beta_{2s+1} = \begin{vmatrix} d_1 & d_2 & d_3 & \cdots & d_{s+1} \\ d_2 & d_3 & d_4 & \cdots & d_{s+2} \\ \vdots & & & & \vdots \\ d_{s+1} & d_{s+2} & d_{s+3} & \cdots & d_{2s+1} \end{vmatrix},$$

$$\beta_{2s} \neq 0, \qquad \beta_{2s+1} \neq 0.$$

The infinite continued fraction found in this manner is called the corresponding continued fraction and the power series is said to be semi-normal if $\beta_{2s} \neq 0$, $\beta_{2s+1} \neq 0$.

2. A new continued fraction. Müller found his continued fraction by converting in the manner of the preceding paragraph

$$I_x(p, q) = \frac{\Gamma(p+q)x^p(1-x)^{q-1}}{\Gamma(p+1)\Gamma(q)} \cdot \left\{1 + \sum_{r=0}^{\infty} \frac{(q-1)(q-2)\cdots(q-r-1)}{(p+1)(p+2)\cdots(p+r+1)} \left(\frac{x}{1-x}\right)^{r+1}\right\}, \quad x < \tfrac{1}{2}. \tag{2.1}$$

We convert

$$I_x(p, q) = \frac{\Gamma(p+q)x^p(1-x)^q}{\Gamma(p+1)\Gamma(q)} \cdot \left\{1 + \sum_{r=0}^{\infty} \frac{(p+q)(p+q+1)\cdots(p+q+r)}{(p+1)(p+2)\cdots(p+r+1)} x^{r+1}\right\}, \quad 0 < x < 1. \tag{2.2}$$

Consequently

$$\beta_1 = \frac{p+q}{p+1}, \qquad \beta_2 = \frac{(p+q)(1-q)}{(p+1)^2(p+2)}, \cdots,$$

$$\beta_{2s+1} = \frac{(p+q)(p+q+1)\cdots(p+q+s-1)(p+q+s)}{(p+s+1)(p+s+2)\cdots(p+2s)(p+2s+1)}\beta_{2s},$$

$$\beta_{2s+2} = \frac{(1-q)(2-q)\cdots(s-q)(s+1-q)(s+1)!}{(p+1)(p+2)\cdots(p+2s+1)(p+2s+2)}\beta_{2s+1},$$

$$c_{2s+1} = -\frac{(p+s)(p+q+s)}{(p+2s)(p+2s+1)}, \qquad c_{2s} = \frac{s(q-s)}{(p+2s-1)(p+2s)},$$

and

$$I_x(p, q) = \frac{\Gamma(p+q)x^p(1-x)^q}{\Gamma(p+1)\Gamma(q)} \left\{\frac{1}{1+}\ \frac{C_1}{1+}\ \frac{C_2}{1+}\ \cdots\right\}, \tag{2.3}$$

where $C_s = c_s x$. By well known theorems due to Van Vleck [12] and Perron [9] p. 347 we find (1.2) converges for $-1 < x < \infty$, and (2.3) converges for $-\infty < x < 1$, and in the neighborhood of zero (2.2) equals (2.3). The region of equivalence of the series and the fraction may be extended by the following argument Let the infinite series be terminated at some arbitrary point which gives the desired accuracy. Then the continued fraction of the corresponding type represents this finite series, is finite and gives the result within the desired accuracy. The new continued fraction may also be derived by use of the hypergeometric series [9] p. 348. A special case of (2.3) was given by Markoff [3], pp 135–41, [11] pp. 53–55, who applied the result only to the binomial distribution. The associated continued fraction provides more rapid convergence than the corresponding continued fraction The associated continued fraction is found by means of the hypergeometric series [9] p. 331, p. 348:

$$I_x(p,q) = \frac{\Gamma(p+q)x^p(1-x)^q}{\Gamma(p+1)\Gamma(q)}\left\{1 + \frac{k_1 x}{1+l_1 x+}\ \frac{k_2 x^2}{1+l_2 x+}\ \frac{k_3 x^2}{1+l_3 x+}\ \cdots\right\}$$

$$k_1 = \frac{p+q}{p+1}, \qquad l_1 = \frac{p+q+1}{p+2},$$

(2.4)

$$k_{s+1} = \frac{s(s-q)(p+s)(p+q+s)}{(p+2s-1)(p+2s)^2(p+2s+1)},$$

$$l_{s+1} = \frac{s(q-s)}{(p+2s)(p+2s+1)} - \frac{(p+s+1)(p+q+s+2)}{(p+2s+2)(p+2s+1)}, \quad s \geqq 1.$$

The disadvantage of (2.4) lies in the unwieldy form of computation. For properties of an associated continued fraction and the corresponding continued fraction in connection with convergence and the Taylor series reference is made to [9] p. 331 and pp. 302–303.

3. Properties of the corresponding continued fraction. Müller and Soper [5], [10], pointed out the inadvisability of integration through the mode $x = \frac{p-1}{p+q-2}$. In such cases we change $I_x(p, q)$ to $I_{1-x}(q, p)$. Müller has shown for his continued fraction that if we do not integrate through the mode (we assume this in the remainder of the paragraph) that convergents 2, 3, 6, 7, etc., will be greater than the true value and the remaining convergents will be less than the true value provided q is an integer. However, if q is not an integer, and is small ($q < 20$), it may happen that all convergents are above the true value. In such cases we may consider whether Müller's continued fraction may apply by estimating the remainder $I(p+s, q-s)$, after s reductions by parts [10].

For the new continued fraction also

$$|C_{2s}| = \left|\frac{s(q-s)}{(p+2s-1)(p+2s)} \cdot \frac{p-1}{p+q-2}\right| < 1,$$

$$|C_{2s+1}| = \left|\frac{(p+s)(p+q+s)(p-1)}{(p+2s)(p+2s+1)(p+q-2)}\right| < 1,$$

and $C_{2s+1} < 0$; $C_{2s} > 0$ unless $s > q$ when $C_{2s} < 0$. If $C_{2s} > 0$ then the convergents 2, 3, 6, 7, 10, 11, etc., will be above the true value and the other convergents will be below the true value If $C_{2s} < 0$, then all convergents will be above the true value. In such cases, since a remainder for the continued fraction has not been found, it seems best to estimate $I_x(p + s, q - s)$ to obtain an idea of the error.

4. $I_x(p + s, q - s)$ and the equivalent continued fraction. Soper [10] has given the remainder after s reductions by raising p. This will furnish an upper bound of the error in the corresponding continued fraction after s convergents. The remainder, when $q - s$ is a negative integer, is approximately

$$I_x(p + s, q - s) = \frac{2 \sin (q - s)\pi \sqrt{\xi(\xi - 1)/2\pi(p + q)}}{\xi - x} \left\{ \left(\frac{x}{\xi}\right)^{\xi} \left(\frac{1 - x}{\xi - 1}\right)^{1-\xi} \right\}^{p+q}, \tag{4.1}$$

$$\text{where } \xi = \frac{p + s}{p + q}.$$

Another approach is to use the equivalent continued fraction, for $s - 1$ convergents of the equivalent continued fraction reproduces exactly s terms of the infinite series. The infinite series and the equivalent continued fraction for the infinite series are alike in all respects except form. By [9] p. 210, we find that the equivalent continued fraction for (2.3) is

$$W_1 = \frac{\gamma_1}{1 + \gamma_1 -} \frac{\gamma_2}{1 + \gamma_2 -} \frac{\gamma_3}{1 + \gamma_3 -} \frac{\gamma_4}{1 + \gamma_4 -} \cdots$$

where

$$\gamma_1 = \frac{p + q}{p + 1} x, \qquad \gamma_2 = \frac{p + q + 1}{p + 2} x, \qquad \gamma_3 = \frac{p + q + 2}{p + 3} x, \cdots \tag{4.2}$$

$$\gamma_r = \frac{p + q + r - 1}{p + r} x,$$

and

$$I_x(p, q) = \frac{\Gamma(p + q)x^p(1 - x)^q}{\Gamma(p + 1)\Gamma(q)} \frac{1}{1 - W_1}.$$

The equivalent continued fraction for Müller's continued fraction is given in [5], p. 292.

5. Numerical illustration. If A_v and B_v represent the numerator and the denominator of the v-th convergent of a continued fraction $\frac{a_1}{b_1+} \frac{a_2}{b_2+} \frac{a_3}{b_3+} \frac{a_4}{b_4+} \cdots$ then'

$$\begin{aligned} A_v &= b_v A_{v-1} + a_v A_{v-2} \\ B_v &= b_v B_{v-1} + a_v B_{v-2}, \end{aligned} \qquad v > 2. \tag{5.1}$$

As an example we calculate $I_{.5}$ (2.5, 1.5), which could not be done by Müller's continued fraction.

Convergent	A	B	A/B
1	1	1	1
2	1	.42857143	2.3333333
3	1.015873016	.44444444	2.2857142
4	.66233767	.29292929	2.2610838
5	.64812966	.28671329	2.2605498
6	.46471308	.20559441	2.2603391
7	.441837914	.195475117	2.2603281
8	.33105492	.14646345	2.2603245
9	.30890766	.13666520	2.2603242
10	.23762461	.10512856	2.2603240
11	.21882154	.096809808	2.2603240

Using the value of the eleventh convergent we have, $I_{.5}$ (2.5, 1.5) = .28779339. Pearson [7], p. 30, gives .2877934 and Soper [10], p. 32 gives .28779341.

6. Discussion of the various methods. Müller's continued fraction encounters difficulties when q is small due to the possible divergence of the series on which it is based. In such cases the new continued fraction works admirably. Where "reduction by parts" [10] is advisable it would seem Müller's results will be better, while if "integration raising p" is preferable, then the new continued fraction would be necessary. The other methods suggested in the past lacked in some cases remainder terms; were in other cases too long; were feasible only in a limited range; or were only approximations. I am particularly indebted to Professor C. C. Craig under whose guidance this study was completed.

REFERENCES

[1] L. A. Aroian, "A study of R. A. Fisher's z Distribution and the related F distribution," unpublished manuscript.

[2] B. H. Camp, "Probability integrals for the point binomial," *Biometrika*, Vol. 16 (1924), pp. 163-171.

[3] A. A. Markoff, *Wahrscheinlichkeitsrechnung*, translated from the second Russian edition by Heinrich Liebmann. Leipzig: B. G. Teubner, 1912.

[4] T. Muir, "New general formulae for the transformation of infinite series into continued fractions," *Roy. Soc. Edinb. Trans.*, Vol. 27, p. 467.

[5] J. H. Müller, "On the application of continued fractions to the evaluation of certain integrals, with special reference to the incomplete Beta function," *Biometrika*, Vol. 22 (1930-31), pp. 284-297.

[6] K. Pearson (Editor), *Tables for Statisticians and Biometricians*, Part II, London: Biometric Laboratory, 1931, pp. cexxv-cexxvi.

[7] K. Pearson, *Tables of the Incomplete Beta-Function*, London: Biometrika Office, 1934.

[8] M. V. and K. Pearson, "On the numerical evaluation of high order Eulerian integrals," *Biometrika*, Vol. 27 (1935), pp. 409-423.

[9] O. PERRON, *Die Lehre von den Kettenbrüchen*, Leipzig: G. Teubner, 1913 (Pages refer to this edition)

[10] H. E. SOPER, *Tracts for Computers No. 7, The Numerical Evaluation of the Incomplete Beta-Function*, London: Cambridge Univ. Press, 1921

[11] J. V USPENSKY, *Introduction to Mathematical Probability*, New York McGraw-Hill Book Co , 1937.

[12] E. B. VAN VLECK, "On the convergence of algebraic continued fractions," *Am Math. Soc. Trans.*, Vol. 5 (1904), pp. 253-262

[13] J. WISHART, "Determination of $\int_0^\theta \cos^{n+1}\theta\, d\theta$ for large values of n, and its application to the probability integral of symmetrical frequency curves," *Biometrika*, Vol. 17 (1925), pp 68-78.

[14] J. WISHART, "Further remarks on a previous paper," *Biometrika*, Vol 17 (1925), pp. 469-471.

[15] J WISHART, "On the approximate quadrature of certain skew curves," *Biometrika*, Vol. 19 (1927), pp. 1-38.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

NOTE ON THE DISTRIBUTION OF NON-CENTRAL t WITH AN APPLICATION

By Cecil C. Craig

University of Michigan

If we adopt the notation recently used by N. L. Johnson and B. L. Welch [1], non-central t is defined by

$$t = \frac{z + \delta}{\sqrt{w}},$$

in which δ is a constant and z and w are independent variables, z being distributed normally about zero with unit variance and w being distributed as χ^2/f in which f is the number of degrees of freedom for χ^2.

In the paper referred to Johnson and Welch discuss some applications of non-central t and give suitable tables calculated from the probability integral of the distribution of this variable. Previously tables of this probability integral for the purpose of calculating the power of the t test had been given by J. Neyman [2] and Neyman and B. Tokarska [3].

It is the purpose of this note to call attention to a series expansion for the probability integral of non-central t which is simple in form and in most cases convenient for direct calculation. As an application of some intrinsic interest this series is used to compute in several numerical cases the power of a test proposed by E. J. G. Pitman [4] based on the randomization principle.

If for convenience we write,

$$\sqrt{w} = \psi, \ (0 \leq \psi \leq \infty),$$

we have for the joint distribution of $z + \delta$ and ψ,

$$df(z + \delta, \psi) = \frac{2(f/2)^{f/2}}{\sqrt{2\pi}\,\Gamma(f/2)} e^{-\frac{1}{2}(f\psi^2 + z^2)} \psi^{f-1}\, d\psi\, dz. \tag{1}$$

From this

$$\begin{aligned} df(t, \psi) &= \frac{2(f/2)^{f/2} e^{-\delta^2/2}}{\sqrt{2\pi}\,\Gamma(f/2)} e^{-\psi^2(f + t^2)/2 + \delta\psi t} \psi^f\, d\psi\, dt \\ &= \frac{2(f/2)^{f/2} e^{-\delta^2/2}}{\sqrt{2\pi}\,\Gamma(f/2)} e^{-\psi^2(f+t^2)/2} \sum_{r=0}^{\infty} \frac{(\delta t)^r}{r!} \psi^{f+r}\, d\psi\, dt, \end{aligned} \tag{2}$$

Now this series can be integrated term by term with respect to ψ over its range and we have,

$$(3) \qquad df(t) = \frac{(f/2)^{f/2} e^{-\delta^2/2}}{\sqrt{2\pi}\,\Gamma(f/2)} \sum_{r=0}^{\infty} \frac{\Gamma[\frac{1}{2}(f+r+1)]}{r!} (\delta t)^r \left(\frac{2}{f+t^2}\right)^{\frac{1}{2}(f+r+1)} dt.$$

This series converges uniformly in any finite interval for t and it may be integrated term by term over the entire range for t or over any part of it In particular, after some reduction, we get,

$$(4) \qquad \begin{aligned} P(0 \leq t \leq t_0 \,|\, f, \delta) &= \int_0^{t_0} df(t) \\ &= \frac{e^{-\delta^2/2}}{2} \sum_{r=0}^{\infty} \frac{(\delta^2/2)^{r/2}}{\Gamma(r/2+1)} I\left((r+1)/2, f/2; \frac{t_0^2}{f+t_0^2}\right), \end{aligned}$$

in which $I\left((r+1)/2, f/2; \frac{t_0^2}{f+t_0^2}\right)$ is the incomplete Beta-function in the notation of Karl Pearson. Often what is wanted is

$$(5) \qquad P(-t_0 \leq t \leq t_0) = e^{-\delta^2/2} \sum_{r=0}^{\infty} \frac{(\delta^2/2)^r}{r!} I\left((r+1)/2, f/2; \frac{t_0^2}{f+t_0^2}\right).$$

Since the incomplete Beta-function is numerically less than unity it is seen that the series (4) or (5) converges rapidly for moderate values of δ such as will ordinarily occur in applications for small samples. The use of Pearson's tables of $I(p, q; x)$ will be convenient since interpolation will be required for only one of the three arguments

As an application let us consider the test proposed by Pitman in the paper referred to above. Two independent samples, $x_1, x_2, \cdots, x_{N_1}$, and $y_1, y_2, \cdots, y_{N_2}$, have been drawn and it is desired in the absence of any information about the two populations from which the samples came to test the hypothesis that they have equal means. A test based on what may be termed the principle of randomization for this situation has been discussed by R. A. Fisher [5] and by E. S Pearson [6] It is as follows: Let the combined sample of $N_1 + N_2$ observations be separated into sets of N_1 observations, $u_1, u_2, \cdots, u_{N_1}$, and N_2 observations, $v_1, v_2, \cdots, v_{N_2}$, in all possible ways. For each such separation let the numerical difference of the means, $|\bar{u} - \bar{v}|$, be the spread. Then for a suitably chosen $\delta > 0$, we will reject the hypothesis of equal means if fewer than $100\alpha\%$ of the ${}_{N_1+N_2}C_{N_1}$ spreads exceed $|\bar{x} - \bar{y}|$, and otherwise not It is clear that this test is fiducially valid independently of the populations actually sampled in the sense that if it be consistently followed for all such samples, the proportion of cases when the hypothesis is rejected when it is true will statistically approach α.

For all but very small samples it is very tedious to calculate the ${}_{N_1+N_2}C_{N_1}$

spreads and Pitman in his discussion shows that for quite moderate values of N_1 and N_2 the quantity,

$$w = \frac{\frac{N_1N_2}{(N_1+N_2)^2}(\bar{u}-\bar{v})^2}{\frac{\Sigma(x-\bar{x})^2+\Sigma(y-\bar{y})^2}{N_1+N_2}+\frac{N_1N_2}{(N_1+N_2)^2}(\bar{u}-\bar{v})^2} = \frac{\zeta^2}{\xi^2+\zeta^2}$$

has a distribution which in all but very exceptional cases is quite well approximated by a $B(\frac{1}{2}, \frac{1}{2}(N_1+N_2-2))$-function. That is, the distribution of w for the ${}_{N_1+N_2}C_{N_1}$ spreads may for practical purposes be found from that of t, by a simple transformation, with N_1+N_2-2 degrees of freedom.

It seems pertinent to make some inquiry into the power of such a test, that is, to make an attempt to learn something about the probability that such a test will fail to reject the hypothesis of equal means when it is in fact false. To do this it is now necessary to specify the populations which have actually been sampled. If we suppose that these populations are normal with equal variances but with unequal means which, with no loss of generality, may be taken to be μ and $-\mu$ respectively, the probability integral of the distribution of non-central t will give our answer.

If we set

$$\frac{t^2}{f+t^2} = \frac{\zeta^2}{\xi^2+\zeta^2},$$

we have

$$t = \sqrt{f}\,\zeta/\xi.$$

Also,

$$\xi^2 = \frac{(N_1-1)s_1^2+(N_2-1)s_2^2}{N_1+N_2-2}\cdot\frac{N_1+N_2-2}{N_1+N_2} = \frac{f}{N_1+N_2}s^2,$$

in which s^2 is the usual estimate of the population variance σ^2 based on $f = N_1+N_2-2$ degrees of freedom. Then

$$t = \frac{\bar{u}-\bar{v}}{s}\sqrt{\frac{N_1N_2}{N_1+N_2}}$$

and this is a central t if $\mu = -\mu = 0$, otherwise it is non-central. In the latter case we write (the test is made on $\bar{x}-\bar{y}$),

$$t = \frac{(\bar{x}-\mu)-(\bar{y}+\mu)+2\mu}{s}\sqrt{\frac{N_1N_2}{N_1+N_2}}$$

$$= \frac{z+\delta}{\psi},$$

in which,

$$z = \frac{(\bar{x} - \mu) - (\bar{y} + \mu)}{\sigma} \sqrt{\frac{N_1 N_2}{N_1 + N_2}},$$

$$\psi = s/\sigma,$$

and

$$\delta = \frac{2\mu}{\sigma} \sqrt{\frac{N_1 N_2}{N_1 + N_2}}.$$

In applying Pitman's test for a given significance level α, one determines whether or not

$$P(w > w_0) \gtrless \alpha,$$

w_0 being the value of w calculated from the sample. This is equivalent to finding

$$P(t^2 > t_0^2),$$

for the proper f, in which

$$\frac{t_0^2}{f + t_0^2} = w_0$$

and this can be found from an ordinary table of the probability integral of the t-distribution.

For a numerical example let $N_1 = N_2 = 10$ so that $f = 18$. If we adopt a 5% significance level we have $t_0^2 = 2.101^2$ for the critical value. Let us suppose that $\mu/\sigma = 0.1$, and calculate the probability that the hypothesis that $\mu = 0$ will be rejected. We have $\delta = 0.1$ and

$$\frac{t_0^2}{f + t_0^2} = 0.1969.$$

Then

$$P(t^2 \leq t_0^2) = e^{-0.1}[I(0.5, 9; 0.1969) + 0.1\, I(1.5, 9; 0.1969) + \frac{0.01}{2!} I(2.5, 9; 0.1969) + \ldots]$$

$$= 0.9292.$$

Four terms of the series were enough to give this result. The probability of rejecting the hypothesis in this case is thus 0.0708.

The following tables show results for $\alpha = 0\ 05$ and 0.01, $\mu/\sigma = 0.1$, 0.2, and 0.5, and $N_1 = N_2 = 10$ and 20.

Values of $P(t^2 > t_0^2)$

$N_1 = N_2 = 10$

α \ μ/σ	0.1	0.2	0.5
0.05	0.0708	0.1355	0.5621
0.01	0.0165	0.0396	0.2940

$N_1 = N_2 = 20$

α \ μ/σ	0.1	0.2	0.5
0.05	0.0947	0.2345	0.8691
0.01	0.0251	0.0862	0.6730

In only one case was it necessary to calculate as many as ten terms of the corresponding series to obtain these values.

REFERENCES

[1] N. L. Johnson and B. L. Welch, "Applications of the non-central t-distribution," *Biometrika*, Vol. 31 (1940), pp. 362–389.

[2] J. Neyman, Statistical problems in agricultural experimentation," *Roy. Stat. Soc. Jour.*, Supplement, Vol. 2 (1935), pp. 127–136.

[3] J. Neyman and B. Tokarska, "Errors of the second kind in testing 'Student's' hypothesis," *Amer. Stat. Assn. Jour.*, Vol. 31 (1936), pp. 318–326.

[4] E. J. G. Pitman, "Significance tests which may be applied to samples from any populations," *Roy. Stat. Soc. Jour.*, Supplement, Vol. 4 (1937), pp. 119–130.

[5] R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, Edinburgh, 1935, Section 21.

[6] E. S. Pearson, "Some aspects of the problem of randomization," *Biometrika*, Vol. 29 (1937), pp. 53–64.

NOTE ON AN APPLICATION OF RUNS TO QUALITY CONTROL CHARTS

By Frederick Mosteller

Princeton University

In the application of statistical methods to quality control work, a customary procedure is to construct a control chart with control limits spaced about the mean such that under conditions of statistical control, or random sampling, the probability of an observation falling outside these limits is a given α (e.g., .05). The occurrence of a point outside these limits is taken as an indication of the presence of assignable causes of variation in the production line. Such a form

of chart has been found to be of particular value in the detection of the presence of assignable causes of variability in the quality of manufactured product. As recently pointed out, however, the statistician may not only help to detect the presence of assignable causes, but also help to discover the causes themselves in the course of further research and development. For this purpose, runs of different kinds and of different lengths have been found useful by industrial statisticians.[1] Quality control engineers have found, at least in research and development work, that a convenient indication of lack of control is the occurrence of long runs of observations whose values lie above or below that of the median of the sample. For example (as will be shown below), at least one succession of 9 or more observations above or below the median in a sample of 40 would be taken as evidence of lack of control at the .05 level; meaning that under conditions of control such a run would occur in approximately 5 per cent of the samples. Since this type of test has been found useful by quality control engineers, it is perhaps desirable to discuss the mathematical basis of such tests of control and provide a brief table for samples of various sizes at the significance levels .05 and .01.

The general distribution theory of runs of k kinds of elements, and in particular that of two kinds has been thoroughly investigated by A. M. Mood.[2] The purpose of this note is to give an application of the general method to quality control.

Let us consider a sample of size $2n$ drawn from a continuous distribution function $f(x)$. These are then arranged in the order in which they were drawn. We now separate the sample into two sets by considering the nth and $(n + 1)$st elements in order of magnitude, then if $x_i \leqq x_n$, x_i will be called an a, and if $x_i \geqq x_{n+1}$, x_i will be called a b. A run of a's will be defined as usual as a succession of a's terminated at each end by the occurrence of a b (with the obvious exceptions where the run includes the first or last element of the sample), and

---

[1] The use of "runs up" and "runs down" as well as runs above and below the arithmetic mean of a sample were briefly described in a paper by W. A. Shewhart, "Contribution of statistics to the science of engineering," before the Bicentennial Celebration of the University of Pennsylvania, September 17, 1940, to be published in the proceedings of that meeting. In a paper, "Mathematical statistics in mass production," presented before the American Mathematical Society in February, 1941, Shewhart discussed some of the advantages of using runs above and below the median and showed how by comparing runs of different types in a given problem it is often possible to fix rather definitely the source of trouble. The present note considers only the frequency of occurrence of "long" runs which are often used by research and development engineers to indicate the presence of assignable causes of variation. The occurrence of more than one such run in a given sequence, if distributed above and below the median value may also constitute valid evidence of the presence of more than one state of statistical control between which the phenomena may oscillate. The interpretation of long runs in this sense, however, is not considered in the present note.

[2] A. M. Mood, "The distribution theory of runs," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 367–392.

runs of b's are defined similarly. A run of a's may conveniently be called a run "below the median," and a run of b's a run "above the median."

We shall use Mood's notation throughout, i.e., r_{1i}, r_{2i}, $(i = 1, 2, \cdots, n)$ are the number of runs of a's and b's respectively of length i, and r_1, r_2 are the total number of runs of a's and b's; $\begin{bmatrix} m \\ m_i \end{bmatrix}$ will indicate a multinomial coefficient, and $\binom{n}{k}$ a binomial coefficient. Also we define

$$F(r_1, r_2) = 0, \qquad |r_1 - r_2| > 1,$$
$$F(r_1, r_2) = 1, \qquad |r_1 - r_2| = 1,$$
$$F(r_1, r_2) = 2, \qquad |r_1 - r_2| = 0.$$

Then the distribution of runs of a's for our case is

$$P(r_{1i}) = \frac{\begin{bmatrix} r_1 \\ r_{1i} \end{bmatrix}\binom{n+1}{r_1}}{\binom{2n}{n}}. \tag{1}$$

We would like to find the probability of at least one run of s or more a's. The coefficient of x^n in

$$[x + x^2 + \cdots + x^{s-1}]^{r_1}, \tag{2}$$

gives the number of ways of partitioning n elements into r_1 partitions such that no partition contains s or more elements, and none is void. Rewriting (2) we have

$$x^{r_1}[(1 - x^{s-1})]^{r_1} \sum_{t=0}^{\infty} \binom{r_1 - 1 + t}{r_1 - 1} x^t,$$

and the coefficient of x^n is just

$$\sum_{j=0}^{r_1} (-1)^j \binom{r_1}{j}\binom{n - j(s-1) - 1}{r_1 - 1}. \tag{3}$$

Then the probability that we desire, of getting at least one run of s or more a's is immediately given by

$$P(r_{1i} \geqq 1, i \geqq s) = \frac{\sum_{r_1=1}^{n-s+1}\left[\binom{n-1}{r_1-1} - \sum_{j=0}^{r_1}(-1)^j\binom{r_1}{j}\binom{n-1-j(s-1)}{r_1-1}\right]\binom{n+1}{r_1}}{\binom{2n}{n}}.$$

Noting that when $j = 0$ in the inner summation we have just the total number of partitions, we get finally

$$(4) \qquad P(r_{1i} \geqq 1,\ i \geqq s) = \frac{\sum_{r_1=1}^{n-s+1} \binom{n+1}{r_1} \sum_{j=1}^{r_1} (-1)^{j+1} \binom{r_1}{j} \binom{n-1-j(s-1)}{r_1-1}}{\binom{2n}{n}}.$$

A similar result of course holds for the b's.

If we desire the probability of getting at least one run of s or more of either a's or b's, we compute the probability of getting no runs of this type and subtract from unity. Expression (3) multiplied by the total number of ways of getting no partitions of s or more b's for a given r_1, and then summed on r_1 gives exactly the number of ways of getting no runs of either a's or b's as great as s. This is

$$(5) \qquad \begin{aligned} A = \sum_{r_1 > n/s} &\left[\sum_{j=0}^{r_1} (-1)^j \binom{r_1}{j} \binom{n-1-j(s-1)}{r_1-1}\right] \\ &\cdot \left[\sum_{r_2=r_1-1}^{r_1+1} F(r_1, r_2) \sum_{i=0}^{r_2} (-1)^i \binom{r_2}{i} \binom{n-1-i(s-1)}{r_2-1}\right], \end{aligned}$$

and the probability desired is

$$(6) \qquad P(r_{1i} \geqq 1 \text{ or } r_{2i} \geqq 1 \text{ or both};\ i \geqq s) = 1 - A \Big/ \binom{2n}{n}.$$

In spite of the complex appearance of A, the sum can be rapidly calculated for any given s, n since the calculations for the sums on i and j need not be duplicated.

In the case of a quality control chart, we set a significance level α for a given n, this determines s the length of run of either type necessary for significance at the level chosen. Suppose we are interested only in runs occurring on *one side* of the median, say above, when $\alpha = .05$, $n = 20$ (i.e., sample size equal to 40). We determine the least value of s which will make the right hand side of equation (4) less than or equal to .05. It turns out that $s = 8$ for this case. This means that under conditions of statistical control, i.e, random sampling, one or more runs of length 8 or more, above the median will occur in approximately 5 per cent of samples of size 40. Naturally an identical result holds when we are considering only runs below the median.

On the other hand, if under the same conditions as given above ($n = 20$, $\alpha = .05$), we are using as our criterion of statistical control the occurrence of runs of length s or greater *either* above or below the median, we must determine the least value of s such that $1 - A/\binom{2n}{n} \leq .05$. This value turns out to be 9. In other words under conditions of statistical control at least one run of at least 9 will occur *either* above or below the median in less than 5 per cent of the cases on the average.

The following table gives smallest lengths of runs for .05 and .01 significance levels for samples of size 10, 20, 30, 40, 50.

	Runs on one side of median		Runs on either side of median	
$2n$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$
10	5	—	5	—
20	7	8	7	8
30	8	9	8	9
40	8	9	9	10
50	8	10	10	11

If there is an odd number of individuals, say $2n + 1$, in the sample, we would choose the value of the median as the dividing line for our sample and treat the data as if there were only $2n$ cases, thus ignoring the median completely.

The following table[3] gives the probabilities of getting at least one run of s or more on *one* side, *either* side, and *each* side of the median for samples of size 10, 20, and 40.

Length of Run (s)	$2n = 10$			$2n = 20$			$2n = 40$		
	One Side	Either Side	Each Side	One Side	Either Side	Each Side	One Side	Either Side	Each Side
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	.976	.992	.960	1.000	1.000	1.000	1.000	1.000	1.000
3	.500	.667	.333	.870	.956	.784	.992	.999	.986
4	.143	.230	.056	.457	.640	.274	.799	.930	.668
5	.024	.040	.008	.178	.293	.064	.450	.650	249
6				.060	.106	.013	.207	.346	.068
7				.017	.032	.002	.087	.158	.016
8				.004	.007	.000	.034	.065	.004
9				.001	.001	.000	.013	.025	.001
10				.000	.000	.000	.005	.009	.000
11							.002	.003	.000
12							.000	.001	.000
13							.000	.000	.000

One method of computing such a table is to use expression (4) to obtain the probabilities on one side, and to use (6) to get probabilities for either side. Then the probabilities for runs on each side may be computed by using the relationship

$$2P \text{ (one side)} - P \text{ (either side)} = P \text{ (each side)}.$$

[3]The author is indebted to Dr. P. S. Olmstead of the Bell Telephone Laboratories for kindly placing this table at his disposal. Dr. Olmstead has pointed out that these probabilities have been found very useful in research and development work.

TEST OF HOMOGENEITY FOR NORMAL POPULATIONS

By G. A. Baker

University of California

1. Introduction. In biological experiments it is often of interest to test whether or not all the subjects can be regarded as coming from the same normal population. If they have not come from the same normal population, usually the most plausible alternative is that the subjects have come from a population which is the combination of two or more normal populations combined in some proportions. The combination of normal populations is a "smooth" alternative to the hypothesis of a single normal population. Such non-homogeneous populations are not the only "smooth" alternatives, of course, but are included among the "smooth" alternatives. If there is reason to believe that the only deviation from a normal population is due to non-homogeneity, then the results of Professor Neyman in his paper [1] are available in studying this problem.

It is desirable not to make any hypotheses about the mean and standard deviation of the sampled population, but to base all computations and tests on the data contained in the sample. Such a viewpoint has been stressed in a previous paper [2] where it was shown that if the sampling is from a normal population, the probability of a deviation from the mean of a first sample of n measured in terms of the standard deviation of the sample is proportional to

$$\frac{dv}{\left(1 + \frac{v^2}{n+1}\right)^{n/2}}. \tag{1.1}$$

The result (1.1) and Neyman's results give rise to a test of homogeneity which is valid for "large" samples. Empirical results show that fairly conclusive evidence of non-homogeneity may be obtained with samples of 100. Samples of 50 or less may be suggestive but rarely decisive.

2. Development of Test. Suppose that a sample of $n + 1$ is drawn from a normal population. It can be regarded as being made up of a first sample of n and a second sample of one. The value of v corresponding to (1.1) can then be computed and its distribution function is (1.1). This partition, of course, can be made in $n + 1$ ways. That is, $n + 1$ values of v are determined from a random sample of $n + 1$ from the original parent. It is true that these values of v are not independent among themselves. The correlation between the values of v, to a first approximation at least, is of the order of $1/n$ and can be neglected if n is "large."

A suitable transformation as discussed in [3], [1] and elsewhere, transforms (1.1) into a rectangular distribution.

If the same computations are made when the sampled population is not

normal, then the resulting values obtained will not be rectangularly distributed. For instance, suppose that the sampled population is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\left(pe^{-\frac{1}{2}(x-m_1)^2/\sigma^2} + qe^{-\frac{1}{2}(x-m_2)^2/\sigma^2}\right) \tag{2.1}$$

we find that the distribution of v based on the first sample of 2 is a very complicated expression involving sums of exponentials and definite integrals of exponentials. To obtain a rectangular distribution if the sampled population is normal, the appropriate transformation to make is

$$\begin{aligned} v &= -\sqrt{3}\cot \pi u \\ dv &= \sqrt{3}\,\pi \csc^2 \pi u\, du. \end{aligned} \tag{2.2}$$

The resulting u-distribution for population (2.1) then is to be compared with the rectangular distribution in the interval from zero to one.

For "large" values of $n + 1$ and for symmetrical non-homogeneous populations composed of two normal components, the u-distribution will be symmetrical about $u = \frac{1}{2}$, less than one near the ends, greater than one for values of u moderately far from $\frac{1}{2}$ and less than one for values of u near $\frac{1}{2}$. A Neyman [1] Ψ_k^2 of order 4 will be necessary to detect a difference of this sort. If the non-homogeneous population of two components is skewed, the u-distribution will still show the same two-humped effect but may be skewed instead of symmetrical. A Neyman Ψ_k^2 of order 4 should still be computed, although Ψ_3^2 may be more significant.

The test then consists of:

(a) computing the $n + 1$ quantities

$$x_i' = \frac{x_i - \bar{x}}{\sqrt{n+1}\,s}, \qquad (i = 1, 2, 3, \cdots, n+1) \tag{2.3}$$

where

$$\begin{aligned} n + 1 &= \text{number in the sample} \\ x_i &= \text{the observed values} \\ x_j &= \text{the observed values except } x_i \end{aligned}$$

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j, \qquad s^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2$$

(b) making the transformation

$$u_i = \int_{-\infty}^{x_i'} \frac{y_0\,dx'}{(1 + x'^2)^{n/2}}, \qquad (i = 1, 2, 3, \cdots, n+1)$$

(c) computing the first four Ψ_k^2's of Neyman's paper [1]

(d) comparing Ψ_k^2 with $\Psi_i^2(k)$ as found from the Incomplete Gamma Function Tables.

If n is large, say $n = 100$, then u is given approximately by the normal probability integral.

If n is small, the values of u are obtained from the Table 25 of Vol. 2 of Pearson's Tables.

Neyman's derivation assumes that $n + 1$ is large and that the u's are independent In this case, if $n + 1$ is large, then the u's are nearly independent, and hence the test is valid. The same procedure can be applied for smaller samples It can not be expected that small differences from normal in the sampled population can be detected with small samples. Empirical results indicate that samples of 100 are necessary for decisive results even when the differences of the sampled population from a normal homogeneous population are large Samples of 50 may be suggestive and in very extreme cases might be decisive.

TABLE I

Empirical Sampling Results

	$k = 1$	$k = 2$	$k = 3$	$k = 4$
Ψ_k^2's for 51 from population A	.0001	.843	2.009	7.464
Ψ_k^2's for 101 from population A	.086	2.403	4.998	12.868
Ψ_k^2's for 101 from population B	.553	.927	7 472	7.485
Ψ_k^2's for 101 from normal	.017	.082	1 288	1.663
$\Psi^2_{(.05)}(k)$'s (Neyman [1])	3.842	5.992	7.815	9.488
$\Psi^2_{(.01)}(k)$'s (Neyman [1])	6.635	9.210	11.345	13.277

It is to be noted that the test makes no assumption about the parameters of the sampled population and does not group the data. The application of the test gives a unique result that does not depend on the judgment of the computer in any respect. In applying the usual chi-square test the computer must choose groupings The choice of groupings as indicated in [5] may change the P-values to very different levels of significance.

3. Empirical results. Samples of 51 and 101 from population A, of 101 from population B, and of 101 from a normal population, were drawn by throwing dice. Populations A and B are given in [4]. Population A is symmetrical and distinctly bimodal. Population B is weakly bimodal and strongly skewed.

For samples from population A it is necessary to compute Ψ_4^2. For samples from population B it may be sufficient to compute Ψ_3^2. The non-homogeneity of the type of population A seems to be somewhat more detectable than of the type of population B. The sample from the normal parent shows close conformity with expectation.

In applying the proposed test for homogeneity the u-values for small independent sets of data can be combined to give a much larger number of u-values.

REFERENCES

[1] J. NEYMAN, "«Smooth Test» for goodness of fit," *Skandinavisk Aktuariedskrift*, (1937), pp. 149-199.

[2] G A BAKER, "The probability that the mean of a second sample will differ from the mean of a first sample by less than a certain multiple of the standard deviation of the first sample," *Annals of Math. Stat.*, Vol. 6 (1935), pp. 197-201

[3] G. A BAKER, "Transformations of bimodal distributions," *Annals of Math. Stat.*, Vol. 1 (1930), pp 334-344.

[4] G. A. BAKER, "The relation between the means and variances, means squared and variance in samples from the combinations of normal populations," *Annals of Math Stat.*, Vol 2 (1931), pp. 333-354.

[5] G A. BAKER, "The significance of the product-moment coefficient of correlation with special reference to the character of the marginal distributions," *Jour. Am. Stat. Assoc*, Vol 25 (1930), pp. 387-396.

A NOTE ON THE POWER OF THE SIGN TEST

BY W. MAC STEWART

University of Wisconsin

1. Introduction. Let us consider a set of N non-zero differences, of which x are positive and $N - x$ are negative; and suppose that the hypothesis tested, H_0, implies, in independent sampling, that x will be distributed about an expected value of $N/2$ in accordance with the binomial $(\frac{1}{2} + \frac{1}{2})^N$. As a quick test of H_0, we may choose to test the hypothesis h_0 that x has the above probability distribution. Defining r to be the smaller of x and $N - x$, the test consists in rejecting h_0 and therefore H_0 whenever $r \leq r(\epsilon, N)$, where $r(\epsilon, N)$ is determined by N and the significance level ϵ.

2. Power of a test. In applying such a test it is of interest to know how frequently it will lead to a rejection of H_0 when H_0 is false and the situation H implies that the probability law of x is $(q + p)^N$, with $p \neq \frac{1}{2}$, thereby indicating an expectation of an unequal number of $+$ and $-$ differences. The probability of rejecting H_0 when H_1 implying $p = p_1$ is true, is termed the *power* of the test of H_0 relative to the alternative H_1.[1] Thus, from the point of view of experimental design the power (P) of the test of H_0 may be considered a function of the alternative hypothesis H_1, the significance level ϵ, and N. As such, the following observations may be noted:

1. The power P_2, for an assumed ϵ, N, and H_2 implying $p = p_2$ is greater than or equal to the power P_1 for ϵ, N and H_1 implying $p = p_1$ where $| p_2 - .50 | > | p_1 - .50 |$.

[1] For an extensive discussion of the power of a test, the reader is referred to J. Neyman and E. S. Pearson, *Statistical Research Memoirs*, Vol. 1 (1936), pp. 3-6.

2. The power P_2 for an assumed H_1, N, and ϵ_2, is greater than or equal to the power P_1 for H_1, N, and ϵ_1, where $\epsilon_2 > \epsilon_1$.

3. The power P_2 for an assumed H_1, ϵ, and N_2 is greater than or equal to the power P_1 for H_1, ϵ, and N_1 where $N_2 > N_1$.

Hence, to increase the power of the test of H_0 relative to a particular H_1, the methods implied in observations 2 and/or 3 may be employed. However, if any increase in an established ϵ is undesirable, the method implied in observation 3 is the alternative.

3. Explanation of table. In the interests of efficiency and economy, two questions then arise: (1) What is the minimum value of N, which, at the significance level ϵ, will give the test of H_0 a power $P > \beta$, relative to a particular alternative hypothesis H_1? (2) For this minimum value of N corresponding to ϵ, what is the maximum value of r? Stated in another manner, the questions are these: "What is the smallest number (min N) of paired samples that must be employed in conjunction with the Sign Test in order that the test of H_0, at the significance level ϵ, shall have a power $P > \beta$ relative to an alternative hypothesis H_1?" (2) If x of these paired samples give rise to a positive difference, and (min $N - x$) a negative difference, and if r be defined as the smaller of x and (min $N - x$); then, what is the maximum value that r may attain and still have the results, at the level ϵ, judged significant?

Table I provides the answers to these questions for the significance level $\epsilon \leqq .05$; and (1) for H_1 implies $p = p_1$ for values of p_1 from .60 to .95 (and thereby from .40 to .05) at intervals of .05; (2) for values of β from .05 to .95 at intervals of .05, and also for $\beta > .99$. For example, assume that a power $P > .80$ relative to the alternative hypothesis H_3 ($p_1 = .70$) is desired. In Table I, the entry appearing in the column headed H_3 ($p_1 = .70$), and in the row $P > .80$ is 49,17—indicating that 49 paired samples are required, of which 17 or less must be of one sign (+ or −) and hence 32 or more must be of the opposite sign in order that the results be significant at the .05 level.

Because of the discreteness of the binomial distribution, it is impossible to maintain the level of significance at .05 or even arbitrarily close to that figure and still hold to the criterion that N shall be at a minimum. For that reason, particularly when min N is small, results significant at .05 according to Table I may be significant at a level ϵ' where ϵ' is considerably less than .05. In general, however, and in particular when min N is large (greater than 50) both the quantities $(.05 - \epsilon')$ and $(P - \beta)$ are small.

4. Illustrative example. Goulden[2] describes a simple experiment in identifying varieties of wheat. In this experiment, a wheat "expert" is presented paired grain samples of two particular varieties of wheat. The object of the

[2] C. H. Goulden, *Methods of Statistical Analysis*, John Wiley and Sons, New York, 1939, p. 2.

experiment is to test the ability of the expert to differentiate between the two varieties by arranging the pairs so that samples of one variety are on the left, say, and samples of the other variety are on the right.

In a problem of this type, it is desirable to have a sufficiently large number, N, of paired samples in order that the following conditions be fulfilled: (1) The probability that a person possessing no discriminating ability pass the test

TABLE I

Minimum number of paired samples and maximum values of related r

$H_0 \sim p_0 = .50$

(5% level of significance, i.e., $\epsilon \leq .05$)

(min N, max r)

Power	H_8 $p_1=.95$	H_7 $p_1=.90$	H_6 $p_1=.85$	H_5 $p_1=.80$	H_4 $p_1=.75$	H_3 $p_1=.70$	H_2 $p_1=.65$	H_1 $p_1=.60$
$0 < P \leq .05$	—	—	—	—	—	—	7,0	6,0
$P > .05$	—	—	—	—	—	7,0	6,0	9,1
$P > .10$	—	—	—	—	7,0	6,0	9,1	17,4
$P > .15$	—	—	—	8,0	6,0	9,1	12,2	25,7
$P > .20$	—	—	—	7,0	10,1	13,2	17,4	37,12
$P > .25$	—	—	8,0	6,0	14,2	12,2	23,6	44,15
$P > .30$	—	—	7,0	11,1	9,1	18,4	25,7	56,20
$P > .35$	—	—	6,0	10,1	12,2	17,4	30,9	65,24
$P > .40$	—	8,0	—	9,1	16,3	20,5	35,11	74,28
$P > .45$	—	7,0	11,1	—	15,3	26,7	42,14	89,35
$P > .50$	—	6,0	10,1	13,2	18,4	25,7	44,15	101,40
$P > .55$	—	—	9,1	12,2	17,4	30,9	51,18	112,45
$P > .60$	—	—	14,2	15,3	20,5	36,11	56,20	125,51
$P > .65$	7,0	11,1	13,2	19,4	23,6	35,11	63,23	143,59
$P > .70$	6,0	10,1	12,2	18,4	25,7	40,13	67,25	158,66
$P > .75$	—	9,1	16,3	17,4	28,8	44,15	79,30	175,74
$P > .80$	—	14,2	15,3	20,5	30,9	49,17	90,35	199,85
$P > .85$	11,1	12,2	18,4	25,7	35,11	56,20	101,40	227,98
$P > .90$	9,1	15,3	17,4	28,8	42,14	65,24	114,46	263,115
$P > .95$	12,2	17,4	23,6	35,11	49,17	79,30	143,59	327,145
$P > .99$	15,3	23,6	30,9	44,15	67,25	110,44	199,85	453,205

through sheer guesswork be less than ϵ; and (2) if past experience has proven that an expert *does* possess the ability to discriminate between the varieties to the extent of placing a proportion, p_1, of the pairs correctly in the long run, then the probability that he will pass the test be P.

Under these conditions, how large an N is required, and for that N, what is the maximum number of pairs that may be incorrectly placed without failing

the test? For alternative hypothesis H_4 ($p_1 = .75$), and for $P > .90$, referring to Table I, it is seen that 42 paired samples must be employed and not more than 14 may be placed incorrectly. Under the same alternative hypothesis, if it be required merely that $P > .50$ (i.e., an expert with an ability of .75 have better than an even chance of passing), then only 18 paired samples are necessary and not more than 4 may be arranged incorrectly.

Thus, before conducting an experiment in which the Sign Test is to be employed, if the experimenter first decides what power the test must have relative to a certain alternative hypothesis; then from the accompanying table he may learn the minimum number of paired samples that are necessary; and the related maximum value of r.

If this procedure is not followed, and an experimenter employs, say 6 paired samples, he may (as can be seen from the table) discover, to his dismay, that "experts" of ability .75 will be unrecognized more than 80% of the time.

MOMENTS OF THE RATIO OF THE MEAN SQUARE SUCCESSIVE DIFFERENCE TO THE MEAN SQUARE DIFFERENCE IN SAMPLES FROM A NORMAL UNIVERSE

By J. D. Williams

Phoenix, Arizona

The following result may have considerable application to trend analysis. The specific problem was proposed to me by R. H. Kent.

Consider a sample $0_n : X_1, X_2, \cdots, X_n$ from a normal population with zero mean and variance σ^2, the variates being arranged in temporal order. We seek the moments of the ratio of δ^2 to S^2, where

$$(n-1)\delta^2 = \sum_{j=1}^{n-1} (X_j - X_{j+1})^2 \tag{1}$$

and

$$nS^2 = \sum_{j=1}^{n} (X_j - \bar{X})^2. \tag{2}$$

Here $\bar{X}$ is the mean of the X_j. In order to simplify the algebra, we will work with quantities A and B defined by

$$\begin{aligned} 2\sigma^2 A &= (n-1)\delta^2, \\ 2\sigma^2 B &= nS^2. \end{aligned} \tag{3}$$

The characteristic function for the joint distribution of A and B is

$$\begin{aligned} \varphi(t_1, t_2) &= E(e^{At_1 + Bt_2}) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \int_{-\infty}^{\infty}\int\cdots\int \exp\left(At_1 + Bt_2 - \frac{1}{2\sigma^2}\sum_{j=1}^{n} X_j^2\right) \prod_{j=1}^{n} dX_j, \end{aligned} \tag{4}$$

where t_1 and t_2 are pure imaginaries. For the method of analysis which will be used here t_1 and t_2 will be considered as real variables. By straight forward methods we have

$$\varphi^{-2}(t_1, t_2) = \begin{vmatrix} a & b & d & \cdot & \cdot & \cdot & \cdot & \cdot & d \\ b & c & b & d & & & & & \cdot \\ d & b & c & b & d & & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & & \cdot & & \cdot & \cdot \\ \cdot & & & & d & b & c & b & d \\ \cdot & & & & & d & b & c & b \\ d & \cdot & \cdot & \cdot & \cdot & \cdot & d & b & a \end{vmatrix} \tag{5}$$

where the determinant is of nth order and its elements are

$$\begin{aligned} a &= 1 - t_1 - (n-1)T \\ b &= t_1 + T \\ c &= 1 - 2t_1 - (n-1)T \\ d &= T = t_2/n. \end{aligned} \tag{6}$$

It can be verified that the determinant has the value

$$\varphi^{-2}(t_1, t_2) = \sum_{j=0}^{n-1} \binom{2n-j-1}{j} (-t_1)^j (1 - t_2)^{n-j-1}, \tag{7}$$

where the symbol $\binom{2n-j-1}{j}$ represents a binomial coefficient. From (7) we find the moments m_j of A/B as follows: Setting

$$t_2 = \sum_{k=1}^{j} t_{2k}, \tag{8}$$

we have

$$\begin{aligned} m_j &= \int_{-\infty}^{0} \int \cdots \int \frac{\partial^j \varphi(t_1, t_2)}{\partial t_1^j}\Big|_{t_1=0} \prod_{k=1}^{j} dt_{2k} \\ &= \frac{2^j \left[\dfrac{d^j}{dt_1^j} \varphi(t_1, 0)\right]_{t_1=0}}{(n-1)(n+1)\cdots(n+2j-3)}. \end{aligned} \tag{9}$$

The result is rather unexpected, for we have established that the moments of A/B are equal to the moments of A divided by the moments of B.

We find the following explicit values for the first few moments m_j:

$$m_0 = 1$$

$$m_1 = 2$$

$$(10)\quad (n-1)(n+1)m_2 = 4(n^2+n-3)$$

$$(n-1)(n+1)(n+3)m_3 = 8(n^3+6n^2+2n-21)$$

$$(n-1)(n+1)(n+3)(n+5)m_4 = 16(n^4+14n^3+53n^2-8n-231).$$

These are valid subject to the restriction $2n-1 \geq j$, because in arriving at the explicit forms we have treated the binomial coefficient $\binom{k}{j}$ as if it were identically equal to $k(k-1)\cdots(k-j+1)/j!$.

From (10) it is easy to pass to the moments of $R = \delta^2/S^2$. For example, we find the mean value and variance of R to be

$$\frac{2n}{n-1}$$

and

$$\frac{4n^2(n-2)}{(n+1)(n-1)^2}$$

respectively.

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

PAGE

On the Integral Equation of Renewal Theory. WILLY FELLER... 243

On the Joint Distribution of the Medians in Samples from a Multivariate Population. A. M. MOOD..... 268

Samples from Two Bivariate Normal Populations. CHUNG TEI HSU..... 270

On Randomness in Ordered Sequences. L. C. YOUNG..... 293

On Certain Likelihood-Ratio Tests Associated with the Exponential Distribution. EDWARD PAULSON..... 301

On the Mathematically Significant Figures in the Solution of Simultaneous Linear Equations. L. B. TUCKERMAN..... 307

On Mechanical Tabulation of Polynomials. J. C. MCPHERSON... 317

On the Probability of the Occurrence of at Least m Events Among n Arbitrary Events. KAI LAI CHUNG..... 328

Notes:

A Note on Sheppard's Correction. CECIL C. CRAIG..... 339

On the Analysis of Variance in Case of Multiple Classifications with Unequal Class Frequencies. ABRAHAM WALD..... 346

The Frequency Distribution of a General Matching Problem. T. N. E. GREVILLE..... 350

On Methods of Solving Normal Equations. PAUL G. HOEL..... 354

Conditions that the Roots of a Polynomial be Less Than Unity in Absolute Value. PAUL A. SAMUELSON..... 360

Values of Mills' Ratio of Area to Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument. ROBERT D. GORDON..... 364

Vol. XII, No. 3 — September, 1941

THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG J. NEYMAN

WITH THE COÖPERATION OF

H. C. Carver
H. Cramér
W. E. Deming
G. Darmois
R. A. Fisher
T. C. Fry
H. Hotelling
R. von Mises
E. S. Pearson
H. L. Rietz
W. A. Shewhart

The Annals of Mathematical Statistics is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the Annals of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the Annals of Mathematical Statistics should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the Annals is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I-IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

Composed and Printed at the
WAVERLY PRESS, Inc.
Baltimore, Md., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

ON THE INTEGRAL EQUATION OF RENEWAL THEORY

By Willy Feller

Brown University

1. **Introduction.** In this paper we consider the behavior of the solutions of the integral equation

$$u(t) = g(t) + \int_0^t u(t-x)f(x)\,dx, \tag{1.1}$$

where $f(t)$ and $g(t)$ are given non-negative functions.[1] This equation appears, under different forms, in population theory, the theory of industrial replacement and in the general theory of self-renewing aggregates, and a great number of papers have been written on the subject.[2] Unfortunately most of this literature is of a heuristic nature so that the precise conditions for the validity of different methods or statements are seldom known This literature is, moreover, abundant in controversies and different conjectures which are sometimes supported or disproved by unnecessarily complicated examples. All this renders an orientation exceedingly difficult, and it may therefore be of interest to give a rigorous presentation of the theory. It will be seen that some of the previously announced results need modifications to become correct.

The existence of a solution $u(t)$ of (1.1) could be deduced directly from a well-known result of Paley and Wiener [21] on general integral equations of form (1.1).[3] However, the case of non-negative functions $f(t)$ and $g(t)$, with which we are here concerned, is much too simple to justify the deep methods used by Paley and Wiener in the general case. Under the present conditions, the existence of a solution can be proved in a simple way using properties of completely monotone functions, and this method has also the distinct advantage of showing some properties of the solutions, which otherwise would have to be proved separately. It will be seen in section 3 that the existence proof becomes most natural if equation (1.1) is slightly generalized. Introducing the summatory functions

$$U(t) = \int_0^t u(x)\,dx, \qquad F(t) = \int_0^t f(x)\,dx, \qquad G(t) = \int_0^t g(x)\,dx, \tag{1.2}$$

[1] For the interpretation of the equation cf section 2.

[2] Lotka's paper [8] contains a bibliography of 74 papers on our subject published before 1939 Yet it is stated that even this list "is not the result of an exhaustive search." At the end of the present paper the reader will find a list of 16 papers on (1.1) which have appeared during the two years since the publication of Lotka's paper.

[3] This has been remarked also by Hadwiger [3].

equation (1 1) can be rewritten in the form

$$U(t) = G(t) + \int_0^t U(t-x)\,dF(x). \tag{1.3}$$

However, (1.3) has a meaning even if $F(t)$ and $G(t)$ are not integrals, provided $F(t)$ is of bounded total variation and the integral is interpreted as a Stieltjes integral. Now for many practical applications (and even for numerical calculations) this generalized form of the integral equation seems to be the most appropriate one and, as a matter of fact, it has sometimes been used in a more or less hidden form (e.g., if all individuals of the parent population are of the same age). Our existence theorem refers to this generalized equation.

We then turn to one of the main problems of the theory, namely the asymptotic behavior of $u(t)$ as $t \to \infty$. It is generally supposed that the solution $u(t)$ "in general" either behaves like an exponential function, or that it approaches in an oscillating manner a finite limit q; the latter case should arise if $\int_0^\infty f(t)\,dt = 1$, thus in particular in the cases of a stable population and of industrial replacement. However, special examples have been constructed to show that this is not always so.[4] In order to simplify the problem and to get more general conditions, we shall first (section 4) consider only the question of convergence in mean, that is to say, we shall study the asymptotic behavior not of $u(t)$ itself but of the mean value $u^*(t) = \frac{1}{t}\int_0^t u(x)\,dx$. The question can be solved completely using only the simplest Tauberian theorems for Laplace integrals. Of course, if $u(t) \to q$ then also $u^*(t) \to q$, but not conversely. The investigation of the precise asymptotic behavior of $u(t)$ is more delicate and requires more refined tools (section 5).

Most of section 6 is devoted to a study of Lotka's well-known method of expanding $u(t)$ into a series of oscillatory components, and it is hoped that this study will help clarify the true nature of this expansion. It will be seen that Lotka's method can be justified (with some necessary modifications) even in some cases for which it was not intended, e.g., if the characteristic equation has multiple or negative real roots, or if it has only a finite number of roots. On the other hand limitations of the method will also become apparent: thus it can occur in special cases that a formal application of the method will lead to a function $u(t)$ which apparently solves the given equation, whereas in reality it is the solution of quite a different equation.

Of course, most of the difficulties mentioned above arise only when the function $f(t)$ has an infinite tail. However, it is known that even computational considerations sometimes require the use of such curves, and, as matter of fact,

[4] Cf. Hadwiger [2] and also Hadwiger, "Zur Berechnung der Erneuerungsfunktion nach einer Formel von V. A Kostitzin," *Mitt Verein schweizerischer Versich.-Math.*, Vol. 34 (1937), pp 37–43.

exponential and Pearsonian curves have been used most frequently in connection with (1.1). It will be seen that even in these special cases customary methods may lead to incorrect results. Besides, our considerations show how much the solution $u(t)$ is influenced by the values of $f(t)$ for $t \to \infty$, and, accordingly, that extreme caution is needed in practice. The last section contains some simple remarks on the practical computation of the solution.

2. **Generalities on equations (1.1) and (1.3).** This section contains a few remarks on the meaning of our integral equation and on an alternative form under which it is encountered in the literature A reader interested only in the abstract theory may pass immediately to section 3

Equation (1.1) can be interpreted in various ways; the most important among them are the following two:

(i) In the theory of industrial replacement (as outlined in particular by Lotka), it is assumed that each individual dropping out is immediately replaced by a new member of zero age. $f(t)$ denotes the density of the probability at the moment of installment that an individual will drop out at age t. The function $g(t)$ is defined by

$$g(t) = \int_0^t \eta(x) f(t - x)\, dx, \tag{2 1}$$

where $\eta(x)$ represents the age distribution of the population at the moment $t = 0$ (so that the number of individuals of an age between x and $x + \delta x$ is $\eta(x)\delta x + o(\delta x)$). Obviously $g(t)$ then represents the rate of dropping out at time t of individuals belonging to the parent population Finally, $u(t)$ denotes the rate of dropping out at time t of individuals of the total population. Now each individual dropping out at time t belongs either to the parent population, or it came to the population by the process of replacement at some moment $t - x$ $(0 < x < t)$, and hence $u(t)$ satisfies (1.1). It is worthwhile to note that in this case

$$\int_0^\infty f(t)\, dt = 1, \tag{2.2}$$

since $f(t)$ represents a density of probability.

(ii) In population theory $u(t)$ measures the rate of female births at time $t > 0$. The function $f(t)$ now represents the reproduction rate of females at age t (that is to say, the average number of female descendants born during $(t, t + \delta t)$ from a female of age t is $f(t)\delta t + o(\delta t)$). If $\eta(x)$ again stands for the age distribution of the parent population at $t = 0$, the function $g(t)$ of (2.1) will obviously measure the rate of production of females at time t by members of the parent population. Thus we are again led to (1.1), with the difference, however, that this time either of the inequalities

$$\int_0^\infty f(t)\, dt \gtreqless 1 \tag{2.3}$$

may occur, the value of this integral shows the tendency of increase or decrease in the total population.

Theoretically speaking, $f(t)$ and $g(t)$ are two arbitrary non-negative functions. It is true that $g(t)$ is connected with $f(t)$ by (2.1); but, since the age distribution $\eta(x)$ is arbitrary, $g(t)$ can also be considered as an arbitrarily prescribed function.

It is hardly necessary to interpret the more general equation (1.3) in detail, it is the straightforward generalization of (1.1) to the case where the increase or decrease of the population is not necessarily a continuous process. This form of the equation is frequently better adapted to practical needs. Indeed, the functions $f(t)$ and $g(t)$ are usually determined from observations, so that only their mean values over some time units (years) are known. In such cases it is sometimes simpler to treat $f(t)$ and $g(t)$ as discontinuous functions, using equation (1.3) instead of (1.1). For some advantages of such a procedure see section 7. It may also be mentioned that the most frequently (if not the only) special case of (1.1) studied is that where $g(t) = f(t)$. Now it is apparent from (2.1) that this means that all members of the parent population are of zero age: in this case, however, there is no continuous age-distribution $\eta(x)$. Instead we have to use a discontinuous function $\eta(x)$ and write (2.1) in the form of a Stieltjes integral. Thus discontinuous functions and Stieltjes integrals present themselves automatically, though in a somewhat disguised form, even in the simplest cases.

At this point a remark may be inserted which will prove useful for a better understanding later on (section 6). In the current literature we are frequently confronted not with (1.1) but with

$$u(t) = \int_0^\infty u(t-x)f(x)\,dx, \tag{2.4}$$

together with the explanation that it is asked to find a solution of (2.4) which reduces, for $t < 0$, to a prescribed function $h(t)$. Now such a function, as is known, exists only under very exceptional conditions, and (2.4) is by no means equivalent to (1.1). The current argument can be boiled down to the following. Suppose first that the function $g(t)$ of (1.1) is given in the special form

$$g(t) = \int_t^\infty h(t-x)f(x)\,dx, \tag{2.5}$$

where $h(x)$ is a non-negative function defined for $x < 0$. Since the solution $u(t)$ of (1.1) has a meaning only for $t > 0$, we are free to *define* that $u(-t) = h(-t)$ for $t > 0$. This arbitrary definition, then, formally reduces (1.1) to (2.4). It should be noted, however, that this function $u(t)$ does not, in general, satisfy (2.4) for $t < 0$, for $h(t)$ was prescribed arbitrarily. Thus we are not, after all, concerned with (2.4) but with (1.1), which form of the equation is, by the way, the more general one for our purposes. If there really existed a solution of (2.4) which reduced to $h(t)$ for $t < 0$, we could of course define $g(t)$ by (2.5) and transform (2.4) into (1.1) by splitting the interval $(0, \infty)$ into the subintervals

$(0, t)$ and (t, ∞). However, as was already mentioned, a solution of the required kind does not exist in general. It will also be seen (section 6) that the true nature of the different methods and the limits of their applicability can be understood only when the considerations are based on the proper equation (1.1) and not on (2.4).

3. Existence of solutions.

THEOREM 1 *Let $F(t)$ and $G(t)$ be two finite non-decreasing functions which are continuous to the right*[5] *Suppose that*

$$F(0) = G(0) = 0, \tag{3.1}$$

and that the Laplace integrals[6]

$$\varphi(s) = \int_0^\infty e^{-st}\, dF(t), \qquad \gamma(s) = \int_0^\infty e^{-st}\, dG(t) \tag{3.2}$$

converge at least for $s > \sigma \geq 0$[7]. *In case that $\lim_{s \to \sigma+0} \varphi(s) > 1$, let $\sigma' > \sigma$ be the root*[8] *of the characteristic equation $\varphi(s) = 1$, in case $\lim_{s \to \sigma+0} \varphi(s) \leq 1$, put $\sigma' = \sigma$.*

Under these conditions there exists for $t > 0$ one and only one finite non-decreasing function $U(t)$ satisfying (1.3). *With this function the Laplace integral*

$$\omega(s) = \int_0^\infty e^{-st}\, dU(t) \tag{3.3}$$

---

[5] It is needless to emphasize that this restriction is imposed only to avoid trivial ambiguities.

[6] The integrals (3.2) should be interpreted as Lebesgue-Stieltjes integrals over *open* intervals; thus

$$\varphi(s) = \lim_{\epsilon \to +0} \int_\epsilon^\infty e^{-st}\, dF(t),$$

which implies that $\varphi(s) \to 0$ as $s \to \infty$. Alternatively it can be supposed that $F(t)$ and $G(t)$ have no discontinuities at $t = 0$ Continuity of $F(t)$ at $t = 0$ means that there is no reproduction at zero age. This assumption is most natural for our problem, but is by no means necessary. In order to investigate the case where $F(t)$ has a saltus $c > 0$ at $t = 0$, one should take the integrals (3.2) over the closed set $[0, \infty]$, so that

$$\varphi(s) = c + \lim_{\epsilon \to +0} \int_\epsilon^\infty e^{-st}\, dF(t).$$

It is readily seen that Theorem 1 and its proof remain valid if $0 < c < 1$. However, if $c > 1$, then (1.3) plainly has no solution $U(t)$ The continuity of $G(t)$ at $t = 0$ is of no importance and is not used in the sequel

[7] The condition is formulated in this general way in view of later applications (cf., e.g., the lemma of section 4) In all cases of practical interest $\sigma = 0$.

[8] $\varphi(s)$ is, of course, monotonic for $s > \sigma$ and tends to zero as $s \to \infty$. In order to ensure the existence of a root of $\varphi(s) = 1$, it is sufficient to suppose that the saltus c of $F(t)$ at $t = 0$ is less than 1 (cf footnote 6)

converges for $s > \sigma'$, *and*

$$\omega(s) = \frac{\gamma(s)}{1 - \varphi(s)}. \tag{3.4}$$

PROOF. A trivial computation shows that for any finite non-decreasing solution $U(t)$ of (1 3) and any $T > 0$ we have

$$\int_0^T e^{-st}\,dU(t) = \int_0^T e^{-st}\,dG(t) + \int_0^T e^{-sx}\,dF(x) \int_0^{T-x} e^{-st}\,dU(t),$$

herein all terms are non-negative and hence by (3.2)

$$\int_0^T e^{-st}\,dU(t) \leq \gamma(s) + \varphi(s) \int_0^T e^{-st}\,dU(t).$$

Now $\varphi(s) < 1$ for $s > \sigma'$, and hence it is seen that the integral (3.3) exists for $s > \sigma'$ and satisfies (3.4). On the other hand it is well-known that the values of $\omega(s)$ for $s > \sigma'$ determine the corresponding function $U(t)$ uniquely, except for an additive constant, at all points of continuity. However, from (1.3) and (3.1) it follows that $U(0) = 0$ and, since by (1.3) $U(t)$ is continuous to the right, the monotone solution $U(t)$ of (1.3), if it exists, is determined uniquely.

To prove the existence of $U(t)$ consider a function $\omega(s)$ defined for $s > \sigma'$ by (3.4). It is clear from (3.2) that $\varphi(s)$ and $\gamma(s)$ are completely monotone functions, that is to say that $\varphi(s)$ and $\gamma(s)$ have, for $s > \sigma$, derivatives of all orders and that $(-1)^n\varphi^{(n)}(s) \geq 0$ and $(-1)^n\gamma^{(n)}(s) \geq 0$. We can therefore differentiate (3.4) any number of times, and it is seen that $\omega^{(n)}(s)$ is continuous for $s > \sigma'$. Now a simple inductive argument shows that $(-1)^n\omega^{(n)}(s)$ is a product of $\{1 - \varphi(s)\}^{-(n+1)}$ by a finite number of completely monotone functions It follows that $(-1)^n\omega^{(n)}(s) \geq 0$, so that $\omega(s)$ is a completely monotone function, at least for $s > \sigma'$. Hence it follows from a well-known theorem of S. Bernstein and D. V. Widder[9] that there exists a non-decreasing function $U(t)$ such that (3.3) holds for $s > \sigma'$. Moreover, this function can obviously be so defined that $U(0) = 0$ and that it is continuous to the right. Using $U(t)$ let us form a new function

$$V(t) = \int_0^t U(t - x)\,dF(x). \tag{3.5}$$

$V(t)$ is clearly non-negative and non-decreasing. It is readily verified (and, of course, well-known) that

$$\psi(s) \equiv \int_0^\infty e^{-st}\,dV(t) = \omega(s)\varphi(s).$$

It follows, therefore, from (3.4) that $\psi(s) = \omega(s) - \gamma(s)$, and this implies, by the

[9] This theorem has been repeatedly proved by several authors; for a recent proof cf. Feller [19]

uniqueness theorem for Laplace transforms, that $V(t) = U(t) - G(t)$. Combining this result with (3 5) it is seen that $U(t)$ is a solution of (1.3).

THEOREM 2. *Suppose that $f(t)$ and $g(t)$ are measurable, non-negative and bounded in every finite interval $0 \leq t \leq T$ Let the integrals*

$$\varphi(s) = \int_0^\infty e^{-st} f(t)\, dt, \qquad \gamma(s) = \int_0^\infty e^{-st} g(t)\, dt \tag{3.6}$$

converge for $s > \sigma$. Then there exists one and only one non-negative solution $u(t)$ of (1.1) which is bounded in every finite interval[10]. *With this function the integral*

$$\omega(s) = \int_0^\infty e^{-st} u(t)\, dt \tag{3 7}$$

converges at least for $s > \sigma'$, where $\sigma' = \sigma$ if $\lim_{s \to \sigma+0} \varphi(s) \leq 1$, and otherwise $\sigma' > \sigma$ is defined as the root of the characteristic equation $\varphi(s) = 1$. For $s > \sigma'$ equation (3.4) holds.

If $f(t)$ is continuous except, perhaps, at a finite number of points then $u(t) - g(t)$ is continuous.

PROOF: Define $F(t)$ and $G(t)$ by (1.2) Under the present conditions these functions satisfy the conditions of Theorem 1, and hence (1.3) has a non-decreasing solution $U(t)$. Consider, then, an arbitrary interval $0 \leq t \leq T$ and suppose that in this interval $f(t) < M$ and $g(t) < M$. If $0 \leq t < t + h \leq T$ we have by (1.3)

$$\begin{aligned}
0 &\leq \frac{1}{h}\{U(t+h) - U(t)\} \\
&\equiv \frac{1}{h}\{G(t+h) - G(t)\} + \frac{1}{h}\int_t^{t+h} U(t+h-x) f(x)\, dx \\
&\qquad + \frac{1}{h}\int_0^t \{U(t+h-x) - U(t-x)\} f(x)\, dx \\
&\leq M + MU(T) + \frac{M}{h}\int_0^t \{U(t+h-x) - U(t-x)\}\, dx \\
&= M + MU(T) + \frac{M}{h}\int_t^{t+h} U(y)\, dy - \frac{M}{h}\int_0^h U(y)\, dy \\
&< M + 2MU(T).
\end{aligned}$$

Thus $U(t)$ has bounded difference ratios and is therefore an integral. The derivative $U'(t)$ exists for almost all t and $0 \leq U'(t) \leq M$. Accordingly we can differentiate (1.3) formally, and since $U(0) = 0$ it follows that $u(t) = U'(t)$ satisfies (1.1) for almost all t. However, changing $u(t)$ on a set of measure zero does not affect the integral in (1.1), and since $g(t)$ is defined for all t it is seen that

[10] Without the assumptions of positiveness and boundedness this theorem reduces to a special case of a theorem by Paley and Wiener [21]; cf. section 1, p. 243

$u(t)$ can be defined, in a unique way, so as to satisfy (1.1) and obtain (1.3). Since the solution of (1.3) was uniquely determined it follows that the solution $u(t)$ is also unique. Obviously equations (3.7) and (3.3) define the same function $\omega(s)$, so that (3.4) holds, and (3.7) converges for $s > \sigma'$.

Finally, if $f(t)$ has only a finite number of jumps, the continuity of $u(t) - g(t)$ becomes evident upon writing (1.1) in the form

$$u(t) - g(t) = \int_0^t u(x)f(t-x)\,dx.$$

4 **Asymptotic properties.** In this section we shall be concerned with the asymptotic behavior as $t \to \infty$ not of $u(t)$ itself but of the mean value $u^*(t) = \frac{1}{t}\int_0^t u(\tau)\,d\tau$. If $u(t)$ tends to the (not necessarily finite) limit C, then obviously also $u^*(t) \to C$, whereas the converse is not necessarily true. For the proof of the theorem we shall need the following obvious but useful

LEMMA: *If $u(t) \geq 0$ is a solution of* (1.1) *and if*

$$u_1(t) = e^{kt}u(t), \qquad f_1(t) = e^{kt}f(t), \qquad g_1(t) = e^{kt}g(t), \tag{4.1}$$

then $u_1(t)$ is a solution of

$$u_1(t) = g_1(t) + \int_0^t u_1(t-x)f_1(x)\,dx.$$

THEOREM 3: *Suppose that using the functions defined in Theorem 2 the integrals*

$$\int_0^\infty f(t)\,dt = a, \qquad \int_0^\infty g(t)\,dt = b, \tag{4.2}$$

are finite.

(i) *In order that*

$$u^*(t) = \frac{1}{t}\int_0^t u(\tau)\,d\tau \to C \tag{4.3}$$

as $t \to \infty$, where C is a positive constant, it is necessary and sufficient that $a = 1$, and that the moment,

$$\int_0^\infty t\,f(t)\,dt = m \tag{4.4}$$

be finite. In this case

$$C = \frac{b}{m}. \tag{4.5}$$

(ii) *If $a < 1$ we have*

$$\int_0^\infty u(t)\,dt = \frac{b}{1-a}. \tag{4.6}$$

(iii) *If $a > 1$ let σ' be the positive root of the characteristic equation $\varphi(s) = 1$* (cf. (3.2)) *and put*[11]

$$\int_0^\infty e^{-\sigma' t} t\, f(t)\, dt = m_1. \tag{4.7}$$

Then

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t e^{-\sigma' \tau} u(\tau)\, d\tau = \frac{b}{m_1}. \tag{4.8}$$

REMARK: The case $a = 1$ corresponds in demography to a population of stationary size. In the theory of industrial replacement only the case $a = 1$ occurs; the moment m is the average lifetime of an individual. The case $a > 1$ corresponds in demography to a population in which the fertility is greater than the mortality. As is seen from (4.8), in this case the mean value of $u(t)$ increases exponentially. It is of special interest to note that in a population with $a < 1$ the integral (4.6) always converges

PROOF: By (4.2) and (3.7)

$$\lim_{s\to+0} \phi(s) = a, \qquad \lim_{s\to+0} \gamma(s) = b. \tag{4.9}$$

If $a < 1$, it follows from (3.4) that $\lim \omega(s) = b/(1 - a)$ is finite. Since $u(t) \geq 0$ this obviously implies that (4.6) holds. This proves (ii).

If $a = 1$ and m is finite, it is readily seen that

$$\lim_{s\to+0} \frac{1 - \varphi(s)}{s} = m,$$

and hence by (3.4)

$$\lim_{s\to+0} s\omega(s) = \lim_{s\to+0} \gamma(s) \lim_{s\to+0} \frac{s}{1 - \varphi(s)} = \frac{b}{m}.$$

By a well-known Tauberian theorem for Laplace integrals of non-negative functions[12] it follows that $u^*(t) \to \frac{b}{m}$. Conversely, if (4.3) holds it is readily seen that[13]

---

[11] (4.2) implies the finiteness of m_1

[12] Cf. e.g. Doetsch [18], p. 208 or 210.

[13] Indeed, if (4.3) holds and if $U(t)$ is defined by (1.2), then there is a $M = M(\epsilon)$ such that $| U(t) - Ct | < M + \epsilon t$. Now

$$\varphi(s) = s \int_0^\infty e^{-st} U(t)\, dt,$$

and hence

$$s\varphi(s) - C = s^2 \int_0^\infty e^{-st} (U(t) - Ct)\, dt,$$

or

$$| s\varphi(s) - C | \leq s^2 \int_0^\infty e^{-st} (M + \epsilon t)\, dt = sM + \epsilon.$$

$$\lim_{s \to +0} s\omega(s) = C,$$

which in turn implies by (3.4) and (4.9) that

$$\lim_{s \to +0} \frac{1 - \varphi(s)}{s} = \frac{b}{C}.$$

This obviously means that the moment (4.4) exists and equals b/C. This proves (i).

Finally case (iii) reduces immediately to (ii) using the above lemma with $k = -\sigma'$. This finishes the proof.

It may be remarked that the finiteness of the integrals (4.2) is by no means necessary for (4.3). This is shown by the following

EXAMPLE: Let

$$f(t) = \frac{1}{2\sqrt{\pi}\, t^{3/2}} e^{-1/4t}, \qquad g(t) = \frac{1}{\sqrt{\pi t}} e^{-1/4t}.$$

It is readily seen that with these functions $a = 1$, but $b = \infty$. Now[14] $\varphi(s) = e^{-\sqrt{s}}$ and $\gamma(s) = e^{-\sqrt{s}}/\sqrt{s}$, so that

$$\omega(s) = \frac{e^{-\sqrt{s}}}{\sqrt{s}\,(1 - e^{-\sqrt{s}})}.$$

Thus $s\omega(s) \to 1$ as $s \to +0$, and hence $u^*(t) \to 1$. In this particular case it can even be shown that the solution $u(t)$ itself tends to 1 as $t \to \infty$.

In practice, however, the integrals (4.2) will always exist, and accordingly we restrict the consideration to this case.

5. **Closer study of asymptotic properties.** In this section we shall deal almost exclusively with the most important special case, namely where

$$\int_0^\infty f(t)\, dt = 1. \tag{5.1}$$

The question has been much discussed whether in this case necessarily $u(t) \to C$ as $t \to \infty$, which statement, if true, would be a refinement of (4.3). Hadwiger [2] has constructed a rather complicated example to show that $u(t)$ does not necessarily approach a limit. Now this can also be seen directly and without any computations. Indeed, if $u(t) \to C$ and if (5.1) holds, then obviously

$$\lim_{t \to \infty} \int_0^t u(t - x) f(x)\, dx = C,$$

and hence it follows from (1.1) that $g(t) \to 0$. In order that $u(t) \to C$ it is therefore

[14] The integrals can be evaluated by elementary methods, and are known, cf. e.g. Doetsch [18], p. 25.

necessary that $g(t) \to 0$, and this proves the assertion. In Hadwiger's example $\limsup g(t) = \infty$, which makes his computations unnecessary.

It can be shown in a similar manner that not even the condition $g(t) \to 0$ is sufficient to ensure that $u(t) \to C$. Some restriction as to the total variation of $f(t)$ seems both necessary and natural (conditions on the existence of derivatives are not sufficient). In the following theorem we shall prove the convergence of $u(t)$ under a condition which is, though not strictly necessary, sufficiently wide to cover all cases of any possible practical interest.

THEOREM 4: *Suppose that with the functions $f(t)$ and $g(t)$ of Theorem 2*

$$\int_0^\infty f(t)\,dt = 1, \qquad \int_0^\infty g(t)\,dt = b < \infty. \tag{5.2}$$

Suppose moreover that there exists an integer $n \geqq 2$ such that the moments

$$m_k = \int_0^\infty t^k f(t)\,dt, \qquad k = 1, 2, \cdots, n, \tag{5.3}$$

are finite, and that the functions $f(t)$, $tf(t)$, $t^2f(t)$, $\cdots$, $t^{n-2}f(t)$ are of bounded total variation over $(0, \infty)$. Suppose finally that

$$\lim_{t\to\infty} t^{n-2}g(t) = 0 \quad \textit{and} \quad \lim_{t\to\infty} t^{n-2}\int_t^\infty g(x)\,dx = 0. \tag{5.4}$$

Then

$$\lim_{t\to\infty} u(t) = \frac{b}{m_1} \tag{5.5}$$

and

$$\lim_{t\to\infty} t^{n-2}\left\{u(t) - \frac{b}{m_1}\right\} = 0. \tag{5.6}$$

REMARK: As it was shown in section 4, the case where $\int_0^\infty f(t)\,dt > 1$ can readily be reduced to the above theorem by applying the lemma of section 4 with $k = \sigma'$, where σ' is the positive root of $\varphi(s) = 1$: it is only necessary to suppose that $e^{-\sigma' t}f(t)$ is of bounded total variation and that $e^{-\sigma' t}g(t) \to 0$. Obviously all moments of $e^{-\sigma' t}f(t)$ exist, so that the above theorem shows that $u_1(t) = e^{-\sigma' t}u(t)$ tends to the finite limit b'/m_1', where

$$b' = \int_0^\infty e^{-\sigma' t}g(t)\,dt, \qquad m_1' = \int_0^\infty e^{-\sigma' t}tf(t)\,dt.$$

Thus in this case and under the above assumptions $u(t) \sim \frac{b'}{m_1'}e^{\sigma' t}$, so that the renewal function increases exponentially as could be expected. If however

$$\int_0^\infty f(t)\,dt < 1,$$

$u(t)$ will in general *not* show an exponential character. If $f(t)$ is of bounded variation and has a finite moment of second order, and if $g(t) \to 0$, then it can be shown that $u(t) \to 0$ However, the lemma of section 4 can be applied only if the integral defining $\varphi(s)$ converges in some negative s-interval containing a value s' such that $\varphi(s') = 1$, and this is in general not the case.

PROOF: The proof of Theorem 4 will be based on a Tauberian theorem due to Haar[15]. With some specializations and obvious changes this theorem can be formulated as follows.

Suppose that $l(t)$ is, for $t \geq 0$, non-negative and continuous, and that the Laplace integral

$$\lambda(s) = \int_0^\infty e^{-st} l(t)\, dt \tag{5.7}$$

converges for $s > 0$. Consider $\lambda(s)$ as a function of the complex variable $s = x + iy$ and suppose that the following conditions are fulfilled:

(i) For $y \neq 0$ the function $\lambda(s)$ (which is always regular for $x > 0$) has continuous boundary values $\lambda(iy)$ as $x \to +0$, for $x \geq 0$ and $y \neq 0$

$$\lambda(s) = \frac{C}{s} + \psi(s), \tag{5.8}$$

where $\psi(iy)$ has finite derivatives $\psi'(iy), \cdots, \psi^{(r)}(iy)$ and $\psi^{(r)}(iy)$ is bounded in every finite interval;

(ii) $$\int_{-\infty}^{+\infty} e^{itv} \lambda(x + iy)\, dy$$

converges for some fixed $x > 0$ uniformly with respect to $t \geq T > 0$;

(iii) $\lambda(x + iy) \to 0$ as $y \to \pm\infty$, uniformly with respect to $x \geq 0$;

(iv) $\lambda'(iy), \lambda''(iy), \cdots, \lambda^{(r)}(iy)$ tend to zero as $y \to \pm\infty$;

(v) The integrals

$$\int_{-\infty}^{y_1} e^{itv} \lambda^{(r)}(iy)\, dy \quad \text{and} \quad \int_{y_2}^{\infty} e^{itv} \lambda^{(r)}(iy)\, dy$$

(where $y_1 < 0$ and $y_2 > 0$ are fixed) converge uniformly with respect to $t \geq T > 0$.

Under these conditions

$$\lim t^r \{l(t) - C\} = 0. \tag{5.9}$$

Now the hypotheses of this theorem are too restrictive to be applied to the solution $u(t)$ of (1.1). We shall therefore replace (1.1) by the more special equation

$$v(t) = h(t) + \int_0^t v(t - x) f(x)\, dx, \tag{5.10}$$

[15] Haar [20] or Doetsch [18], p. 269.

where

$$h(t) = \int_0^t f(t-x)f(x)\,dx. \tag{5.11}$$

Plainly Theorem 2 can be applied to (5.10). It is also plain that $h(t)$ is bounded and non-negative and that (by (5 1))

$$\int_0^\infty h(t)\,dt = 1, \tag{5.12}$$

$$\chi(s) \equiv \int_0^\infty e^{-st}h(t)\,dt = \varphi^2(s). \tag{5.13}$$

Accordingly we have by Theorem 2

$$\zeta(s) \equiv \int_0^\infty e^{-st}v(t)\,dt = \frac{\varphi^2(s)}{1-\varphi(s)}. \tag{5.14}$$

We shall first verify that $\zeta(s)$ satisfies the conditions of Haar's theorem with $r = n - 2$. For this purpose we write

$$f(t) = f_1(t) - f_2(t), \tag{5.15}$$

where $f_1(t)$ and $f_2(t)$ are non-decreasing and non-negative functions which are, by assumption, bounded.

$$0 \leq f_1(t) < M, \qquad 0 \leq f_2(t) < M. \tag{5.16}$$

(a) We show that $v(t)$ is continuous. Now by Theorem 2 the solution $v(t)$ of (5.10) is certainly continuous if $h(t)$ is continuous; however, that $h(t)$ is continuous follows directly from (5.11) and the fact that the functions

$$\int_0^t f_1(t-x)f(x)\,dx \quad \text{and} \quad \int_0^t f_2(t-x)f(x)\,dx$$

are continuous.

(b) In view of (5.1) the function $\varphi(s)$ exists for $x = \Re(s) \geq 0$. Obviously $|\varphi(x+iy)| < 1$ for $x > 0$. Now

$$\begin{aligned} 1 - \varphi(iy) &= \int_0^\infty (1 - e^{-iyt})f(t)\,dt \\ &= \int_0^\infty (1 - \cos yt)f(t)\,dt + i\int_0^\infty \sin yt \cdot f(t)\,dt, \end{aligned}$$

and, since $1 - \cos yt \geq 0$ and $f(t) \geq 0$, the equality $\varphi(iy) = 1$ for $y \neq 0$ would imply that $f(t) = 0$ except on a set of measure zero. It is therefore seen that $\varphi(x+iy) \neq 1$ for all $x > 0$ and for $x = 0$, $y \neq 0$.

It follows furthermore from (5.3) that for $k = 1, \cdots, n$ and $x \geq 0$ the derivatives

$$\varphi^{(k)}(s) = \int_0^\infty (-t)^k e^{-st}f(t)\,dt$$

exist and that

$$\lim_{x \to +0} \varphi^{(k)}(x + iy) = \varphi^{(k)}(iy).$$

Finally, it is readily seen that in the neighborhood of $y = 0$ we have

$$\varphi(iy) = \int_0^\infty e^{-iyt} f(t)\, dt$$

$$= 1 - m_1 iy + \frac{m_2}{2}(iy)^2 - + \cdots \tag{5.17}$$

$$+ (-1)^{n-1} \frac{m_{n-1}}{(n-1)!}(iy)^{n-1} + O(|y|^n).$$

(c) From what was said under (b) it follows by (5.14) that $\zeta(s)$ is regular for $x > 0$, and that $\zeta(s), \zeta'(s), \cdots, \zeta^{(n)}(s)$ approach continuous boundary values as $s = x + iy$ approaches a point of the imaginary axis other than the origin. Now put

$$\psi(s) = \frac{\varphi^2(s)}{1 - \varphi(s)} - \frac{1}{m_1 s}, \tag{5.18}$$

so that by (5.14)

$$\zeta(s) = \frac{1}{m_1 s} + \psi(s). \tag{5.19}$$

For $x > 0$ and $x = 0$, $y \neq 0$ the function $\psi(x + iy)$ is obviously continuous; the derivatives $\psi'(iy), \cdots, \psi^{(n)}(iy)$ exist. To investigate the behavior of $\psi(iy)$ in the neighborhood of $y = 0$ put

$$P(y) = m_1 - \frac{m_2}{2}(iy) + - \cdots - (-1)^{n-1} \frac{m_{n-1}}{(n-1)!}(iy)^{n-2}. \tag{5.20}$$

By (5.17), (5.18) and (5.20)

$$\psi(iy) = \left[\frac{\{1 - iyP(y)\}^2}{P(y)} - \frac{1}{m_1}\right]\frac{1}{iy} + O(|y|^{n-2}). \tag{5.21}$$

Now the expression in brackets represents an analytic function of y which vanishes at $y = 0$. Hence $\psi(iy) = \mathfrak{P}(y) + O(|y|^{n-2})$, where $\mathfrak{P}(y)$ denotes a power series. It follows that the derivatives $\psi'(iy), \cdots, \psi^{(n-2)}(iy)$ exist for all real y (including $y = 0$) and are bounded for sufficiently small $|y|$: since they are continuous functions they are bounded in every finite interval.

(d). Next we show that there exists a constant $A > 0$ such that for sufficiently large $|y|$

$$|\varphi(x + iy)| < \frac{A}{|y|} \tag{5.22}$$

uniformly in $x \geq 0$. By (5.15)

$$\varphi(s) = \int_0^\infty \{\cos yt - i \sin yt\} e^{-xt} \{f_1(t) - f_2(t)\}\, dt. \tag{5.23}$$

Now $f_1(t)$ is non-decreasing and accordingly by the second mean-value theorem we have for any $T > 0$ and y

$$\int_0^T \cos yt \cdot f_1(t)\, dt = f_1(T) \int_\tau^T \cos yt\, dt = f_1(T) \frac{\sin Ty - \sin \tau y}{y},$$

where τ is some value between 0 and T (depending, of course, on y; at points of discontinuity, $f_1(T)$ should be replaced by $\lim_{t \to T-0} f_1(t)$). Hence by (5.16)

$$\left| \int_0^\infty \cos yt \cdot e^{-xt} \cdot f_1(t)\, dt \right| < \frac{2M}{|y|}.$$

Treating the other terms in (5.23) in a like manner, (5.22) follows.

Combining (5.22) with (5.14) it is seen that for sufficiently large $|y|$

$$|\zeta(s)| < \frac{2A^2}{y^2}$$

uniformly in $x \geq 0$. This shows that the assumptions (ii) and (iii) of Haar's theorem are satisfied for $\lambda(s) = \zeta(s)$. In order to prove that also conditions (iv) and (v) are satisfied it suffices to notice that the proof of (5.22) used only the fact that $f(t)$ is of bounded total variation. Now $\varphi^{(k)}(s)$ is the Laplace transform of $(-t)^k f(t)$, and, since $t^k f(t)$ is of bounded total variation for $k \leq n - 2$, it follows that

$$|\varphi^{(k)}(s)| = O(|y|^{-1}), \qquad k = 1, 2, \cdots, n - 2,$$

for sufficiently large $|y|$, uniformly in $x \geq 0$. Differentiating (5.14) k times it is also seen that

$$|\zeta^{(k)}(s)| = O(|y|^{-2}), \qquad k = 1, 2, \cdots, n - 2,$$

as $y \to +\infty$, uniformly with respect to $x \geq 0$.

This enumeration shows that $v(s) = l(t)$ and $\lambda(s) = \zeta(s)$ satisfy all hypotheses of Haar's theorem with $r = n - 2$ and $C = 1/m_1$. Hence

$$\lim_{t \to \infty} t^{k-2} \left\{ v(t) - \frac{1}{m_1} \right\} = 0. \tag{5.24}$$

Returning now to (5.14) we get

$$\omega(s) = \gamma(s) + \gamma(s)\varphi(s) + \gamma(s)\zeta(s),$$

or, by the uniqueness property of Laplace integrals,

$$\begin{aligned} u(t) &= g(t) + \int_0^t g(x) f(t - x)\, dx + \int_0^t g(x) v(t - x)\, dx \\ &= g(t) + u_1(t) + u_2(t) \end{aligned} \tag{5.25}$$

(which relation can also be checked directly using (5.10)). Let us begin with the last term. We have by (5.2)

$$u_2(t) - \frac{b}{m_1} \equiv \int_0^t g(t-x)\left\{v(x) - \frac{1}{m_1}\right\}dx,$$

and hence

$$t^{n-2}\left|u_2(t) - \frac{b}{m_1}\right| \leq 2^{n-2}\int_{t/2}^t g(t-x)x^{n-2}\left|v(x) - \frac{1}{m_1}\right|dx + t^{n-2}\int_{t/2}^t g(y)\left|v(t-y) - \frac{1}{m_1}\right|dy.$$

If t is sufficiently large we have by (5.24) in the first integral $x^{n-2}\left|v(x) - \frac{1}{m_1}\right| < \epsilon$. In the second integral $v(t-y) - \frac{1}{m_1}$ is bounded, and hence by (5.4)

$$\lim_{t\to\infty} t^{n-2}\left|u_2(t) - \frac{b}{m_1}\right| = 0.$$

The same argument applies (even with some simplifications) also to the second term in (5.24); it follows that

$$\lim t^{n-2}u_1(t) = 0,$$

whilst $t^{n-2}g(t) \to 0$ by assumption (5.4). Now the assertion (5.6) of our theorem follows in view of (5.25) if the last three relationships are added. This finishes the proof of Theorem 4.

It seems that the solution $u(t)$ is generally supposed to oscillate around its limit b/m_1 as $t \to \infty$. It goes without saying that such a behavior is a priori more likely than a monotone character. It should, however, be noticed that there is no reason whatsoever to suppose that $u(t)$ *always* oscillates around its limit. Again no computation is necessary to see this, as shown by the following

EXAMPLE: Differentiating (1.1) formally we get

$$u'(t) = g'(t) + g(0)f(t) + \int_0^t u'(t-x)f(x)\,dx,$$

which shows that, if $g(t)$ and $f(t)$ are sufficiently regular, $u'(t)$ satisfies an integral equation of the same type as $u(t)$. Thus if

$$g'(t) + g(0)f(t) \geq 0$$

for all t, we shall have $u'(t) \geq 0$, and $u(t)$ is a monotone function. In particular, if $g'(t) + g(0)f(t) = 0$, then $u'(t) = 0$ and $u(t) = \text{const}$. For example, let $f(t) = g(t) = e^{-t}$. Then $\varphi(s) = \gamma(s) = 1/(s+1)$ and hence $\omega(s) = 1/s$, which is the Laplace transform of $u(t) = 1$. It is also seen directly that $u(t) \equiv 1$ is the solution. We have however the following

THEOREM 5[16]. *If the functions $f(t)$ and $g(t)$ of Theorem 4 vanish identically for $t \geq T > 0$, then the solution $u(t)$ of (1.1) oscillates around its limit b/m as $t \to \infty$.*

[16] Under some slight additional hypotheses and with quite different methods this theorem was proved by Richter [16].

PROOF: For $t \geq T$ equation (1.1) reduces to

$$u(t) = \int_{t-T}^{t} u(t-x)f(x)\,dx,$$

and since $\int_{t-T}^{t} f(x)\,dx = 1$ it follows that the maxima of $u(t)$ in the intervals $nT < t < (n+1)T$ form, for sufficiently large integers n, a non-increasing sequence. Similarly the corresponding minima do not decrease Since $u(t) \to b/m_1$, by Theorem 4, it follows that the minima do not exceed b/m_1 and the maxima are not smaller than b/m_1.

6. **On Lotka's method.** Probably the most widely used method for treating equation (1.1) in connection with problems of the renewal theory is Lotka's method. As a matter of fact this method consists of two independent parts. The first step aims at obtaining the exact solution of (1.1) in the form of a series of exponential terms (this is achieved by an adaptation of a method which was used by P. Herz and Herglotz for other purposes The second part of Lotka's theory consists of devices for a convenient approximative computation of the first few terms of the series While restricting ourselves formally to Lotka's theory, it will be seen that some of the following remarks apply equally to other methods.

Lotka's method rests essentially on the fundamental assumption that the characteristic equation

$$\varphi(s) = 1 \tag{6.1}$$

has infinitely many distinct simple[17] roots $s_0, s_1, \cdots$, and that the solution $u(t)$ of (1.1) can be expanded into a series

$$u(t) = \sum_k A_k e^{s_k t} \tag{6.2}$$

where the A_k are complex constants. The argument usually rests on an assumed completeness-property of the roots Thus, starting from (2.4) it is required that (6.2) reduces to $h(t)$ for $t < 0$; in other words, that an arbitrarily prescribed function $h(x)$ be, for $x < 0$, respresentable in the form

$$h(x) = \sum_k A_k e^{s_k x} \qquad (x < 0). \tag{6.3}$$

In practice we are, of course, usually not concerned with $h(t)$ but with $g(t)$ (cf. (2.5)), and according to Lotka's theory the coefficients A_k of the solution (6.2) of (1.1) can be computed directly from $g(t)$ in a way similar to the computation of the Fourier coefficients.

Lotka's method is known to lead to correct results in many cases and also to

[17] Hadwiger [3] objected to the assumption that all roots of (6.1) be simple The modifications which are necessary to cover the case of multiple roots also will be indicated below

have distinct computational merits. On the other hand it seems to require a safer justification, since its fundamental assumptions are rarely realized. Thus clearly an arbitrary function $h(x)$ cannot be represented in the form (6.3): to see this it suffices to note that (6.1) frequently has only a finite number of roots (cf. also below) It should also be noted that, the series (6.3) having regularity properties as are assumed in Lotka's theory, any function representable in the form (6.3) is necessarily a solution of the integral equation (2.4), whereas the theory requires us to construct a solution $u(t)$ which reduces to an *arbitrarily* prescribed function $h(t)$ for $t < 0$, (which frequently is an empirical function, determined by observations). Nevertheless, it is possible to give sound foundations to Lotka's method so that it can be used (with some essential limitations and modifications) sometimes even in cases for which it originally was not intended. For this purpose it turns out to be necessary that all considerations be based on the more general equation (1.1), instead of (2.4) (cf. also section 2).

Before proceeding it is necessary to make clear *what is really meant by a root of* (6.1). The function $\varphi(s)$ is defined by (3.2), and the integral will in general converge only for s-values situated in the half-plane $\Re(s) > \sigma$. Usually only roots situated in this half-plane are considered[18]. It is also argued that $\varphi(s)$ is, for real s, a monotone function, so that (6.1) has at most one real root: accordingly the terms of (6.2) are called "oscillatory components." However, the function $\varphi(s)$ can usually be defined by analytic continuation even outside the half-plane $\Re(s) > \sigma$, and, if this is done, (6.1) will in general also have roots in the half-plane $\Re(s) < \sigma$. It will be seen in the sequel that these roots play exactly the same role for the solution $u(t)$ as the other ones, and that the applicability of Lotka's method depends on the behavior of $\varphi(s)$ in the entire complex s-plane. It may be of interest to quote an example where (6.1) has infinitely many real and no other roots.

EXAMPLE[19]: Let

$$f(t) = \frac{1}{2\sqrt{\pi}\, t^{3/2}} e^{-1/4t}, \qquad t > 0; \tag{6.4}$$

---

[18] This was stated in particular by Hadwiger [3] and Hadwiger and Ruchti [6]; accordingly the results of the latter paper (obtained by methods quite different from Lotka's) need some modifications.

[19] Cf. the example at the end of section 4. A function closely related to (6.4) plays an important role in two recent papers by Hadwiger [4] and [5]. Hadwiger's conclusion, if it could be justified, would fundamentally change the aspect of the whole theory. The conclusion reached by Hadwiger seems to be that for any biological population the reproduction function should be of the form $u(t) = \Sigma u_n(t)$, where $u_n(t)$ represents the contribution of the nth generation and

$$(*) \qquad u_n(t) = \frac{an}{\sqrt{\pi}\, t^{3/2}} e^{-At + Can - n^2a^2/t}.$$

Here a, A and C are constants Clearly (*) is a generalization of (6 4) Now his conclusion is based on the arbitrary assumption that $u_n(t)$ should be of the form $u_n(t) = \psi(x, na)$

It is easily seen that $\varphi(s) = e^{-\sqrt{s}}$. The integral (3.2) converges only for $\Re(s) \geq 0$, but $\varphi(s)$ is defined as a two-valued function in the entire s-plane. The roots of (6.1) are obviously $s_k = -4\,k^2\pi^2$, so that all of them are real and simple. If $g(t) = f(t)$, we get by (3.4)

$$\omega(s) = \frac{e^{-\sqrt{s}}}{1 - e^{-\sqrt{s}}} = \sum_1^\infty e^{-n\sqrt{s}}, \qquad s \text{ real}, > 0.$$

Now $e^{-n\sqrt{s}}$ is the Laplace transform of $\frac{n}{2\sqrt{\pi}\,t^{3/2}} e^{-n^2/4t}$, and hence it is readily seen that the solution $u(t)$ can be written in the form

$$u(t) = \frac{1}{2\sqrt{\pi}\,t^{3/2}} \sum_1^\infty n e^{-n^2/4t}; \tag{6.5}$$

of course, this expansion is not of form (6.2) and shows no oscillatory character.

From now on we shall consistently denote by $\varphi(s)$ the function defined by the integral (3 4) and by the usual process of analytic continuation; accordingly we shall take into consideration *all* roots of (6 1). The main limitation of Lotka's theory can then be formulated in the following way: Lotka's method depends only on the function $g(t)$ and on the roots of (6.1) Now two different functions $f(t)$ can lead to characteristic equations having the same roots Lotka's method would be applicable to both only if the corresponding two integral equations (1.1) had the same solution $u(t)$. This, however, is not necessarily the case. Thus, if Lotka's method is applied, and if all computations are correctly performed, and if the resulting series for $u(t)$ converges uniformly, there is no possibility of telling which equation is really satisfied by the resulting $u(t)$· it can happen that one has unwittingly solved some unknown equation of type (1.1) which, by chance, leads to a characteristic equation having the same roots as the characteristic equation of the integral equation with which one was really concerned. Indeed this happens in the following example which is familiar in connection with our problem. It is illustrative also for other purposes: thus it shows not only limitations of Lotka's method, but also that this method can be modified so as to become applicable in some cases where the characteristic equation has only a finite number of roots.

---

where $\psi(x, a)$ is independent of n. To my mind Hadwiger's result shows only the impractibility of this axiom. However, Hadwiger's result is not correct even under his assumption. Indeed, he derives for $\psi(x, a)$ the functional equation

$$\psi(x, a + b) = \int_0^t \psi(x - \xi, a)\psi(\xi, b)\, d\xi, \tag{**}$$

which is well-known from the theory of stochastic processes. Now Hadwiger merely verifies the known result that (*) leads to a solution of (**) However, (**) has infinitely many other solutions (it is possible to write down expressions for their Laplace transforms, although it is difficult to express the solutions themselves explicitly) This, of course, renders Hadwiger's result illusory.

EXAMPLE: *Pearson type III-curves.*[20] Consider the integral equation (1.1) in the following two cases:

$$\text{(I)} \qquad f(t) = g(t) = f_{\mathrm{I}}(t) = \frac{1}{\Gamma(\frac{3}{2})}\, t^{1/2} e^{-t}$$

and

$$\text{(II)} \qquad f(t) = g(t) = f_{\mathrm{II}}(t) = \tfrac{1}{2} t^2 e^{-t}.$$

It is readily seen (and well known) that the corresponding Laplace transforms are

$$\text{(I)} \qquad \varphi_{\mathrm{I}}(s) = \frac{1}{(s+1)^{3/2}}$$

and

$$\text{(II)} \qquad \varphi_{\mathrm{II}}(s) = \frac{1}{(s+1)^3},$$

respectively. Thus in both cases the characteristic equation has the same roots, namely

$$s_1 = 0, \qquad s_{2,3} = -\frac{3}{2} \pm \frac{i}{2}\sqrt{3},$$

of which only the first one lies in the half-plane of convergence of the integral (3.4). Lotka's method is not applicable since there are only three roots. However, in the second case, an expansion of type (6.2) is possible. Indeed, we have by (3.4)

$$\omega_{\mathrm{II}}(s) = \frac{\varphi_{\mathrm{II}}(s)}{1 - \varphi_{\mathrm{II}}(s)} = \frac{1}{s^3 + 3s^2 + 3s}$$

$$= \frac{1}{3s} - \frac{\frac{1}{6} - \frac{i}{2\sqrt{3}}}{s + \frac{3}{2} - \frac{i}{2}\sqrt{3}} - \frac{\frac{1}{6} + \frac{i}{2\sqrt{3}}}{s + \frac{3}{2} + \frac{i}{2}\sqrt{3}};$$

now $1/(s+a)$ is the Laplace transform of e^{-at}, and hence we obtain the solution $u(t)$ in the form

$$u_{\mathrm{II}}(t) = \frac{1}{3} - \left(\frac{1}{6} - \frac{i}{2\sqrt{3}}\right) e^{\frac{1}{2}(-3+i\sqrt{3})t} - \left(\frac{1}{6} + \frac{i}{2\sqrt{3}}\right) e^{\frac{1}{2}(-3-i\sqrt{3})t}$$

$$= \frac{1}{3} - \frac{1}{3} e^{-3t/2} \cos\frac{\sqrt{3}}{2}\, t - \frac{1}{\sqrt{3}} e^{-3t/2} \sin\frac{\sqrt{3}}{2}\, t,$$

[20] General Pearson curves have been investigated recently in connection with (1.1) by Brown [1], Hadwiger and Ruchti [6] and Rhodes [15]. Hadwiger and Ruchti use a method of their own, but they are also led to the study of the characteristic equation (6.1) in a slightly disguised form: their result needs a modification since they arbitrarily drop the roots lying in the halfplane of divergence of the integral $\varphi(s)$.

which is an expansion of type (6.2). In the first of the above examples we get for real positive s

$$\omega_I(s) = \frac{\varphi_I(s)}{1 - \varphi_I(s)} = \sum_{n=1}^{\infty} \frac{1}{(s+1)^{3n/2}},$$

and it is readily seen that this is the Laplace transform of the solution

$$u_I(t) = e^{-t} \sum_{n=1}^{\infty} \frac{1}{\Gamma(3n/2)} t^{3(n-2)/2}.$$

The series is convergent for $t > 0$, but obviously this solution cannot be represented in a form similar to (6.2).

A similar remark applies to the general Pearson-type III curve

$$f(t) = At^{\beta} e^{-\alpha t},$$

where A, α, β are positive constants; the corresponding Laplace transform is

$$\varphi(s) = A\Gamma(\beta + 1) \frac{1}{(s+\alpha)^{\beta+1}}.$$

These preparatory remarks enable us to formulate rigorous conditions for the existence of an expansion of type (6.2). The following theorem shows the limits of Lotka's method, but at the same time it also represents an extension of it. In the formulation of the theorem we have considered only the case of absolute convergence of (6.2). This was done to avoid complications lacking any practical significance whatsoever. The conditions can, of course, be relaxed along customary lines.

THEOREM 6: *In order that the solution $u(t)$ of Theorem 2 be representable in form* (6.2), *where the series converges absolutely for $t \geq 0$ and where the s_k denote the roots of the characteristic equation*[21] (6.1), *it is necessary and sufficient that the Laplace transform $\omega(s)$ admit an expansion*

$$\omega(s) \equiv \frac{\gamma(s)}{1 - \varphi(s)} = \sum \frac{A_k}{s - s_k} \tag{6.6}$$

and that $\Sigma | A_k |$ converges absolutely. The coefficients A_k are determined by

$$A_k = -\frac{\gamma'(s_k)}{\varphi(s_k)}. \tag{6.7}$$

In particular, it is necessary that $\omega(s)$ be a one-valued function.[22]

PROOF: All roots s_k of (6.1) satisfy the inequality $\Re(s_k) \leq \sigma'$, where σ' was defined in Theorem 2. It is therefore readily seen that in case $\Sigma | A_k |$ converges, the Laplace transform of (6.2) can be computed for sufficiently large

[21] The number of roots may be finite or infinite. It should also be noted that it is not required that $s_k \to \infty$. If the s_k have a point of accumulation, $\omega(s)$ will have an essential singularity. That this actually can happen can be shown by examples.

[22] This was not so in our example I.

positive s-values by termwise integration so that (6.6) certainly holds for sufficiently large positive s. Now with $\Sigma \mid A_k \mid$ converging, (6.6) defines $\omega(s)$ uniquely for all complex s (with singularities at the points s_k and the points of accumulation of s_k, if any). Since the analytic continuation is unique, it follows that (6.6) holds for all s. The series $\Sigma \mid A_k \mid$ must, of course, converge if (6.2) is to converge absolutely for $t = 0$, and this proves the necessity of our condition. Conversely, if $\omega(s) = \dfrac{\gamma(s)}{1 - \varphi(s)}$ is given by (6.6), and if $\Sigma \mid A_k \mid$ converges, then $\omega(s)$ is the Laplace transform of a function $u(t)$ defined by (6.2). Since the Laplace transform is unique, $u(t)$ is the solution of (1.1) by Theorem 2. The series (6.2) converges absolutely for $t \geq 0$ since $\mid A_k e^{s_k t} \mid \leq \mid A_k \mid e^{\sigma' t}$. Finally (6.7) follows directly from (6.6).

It is interesting to compare (6.7) with formulas (50) and (56) of Lotka's paper [8]. Lotka considers the special case $g(t) = f(t)$; in this case $\gamma(s_k) = \varphi(s_k) = 1$, and (6.7) reduces to $A_k = -\dfrac{1}{\varphi'(s_k)}$. If s_k lies in the domain of convergence of the integral $\varphi(s) = \displaystyle\int_0^\infty e^{-st} f(t)\, dt$, that is, if $\Re(s_k) \geq \sigma$ then

$$\frac{1}{A_k} = \int_0^\infty e^{-s_k t} t f(t)\, dt, \tag{6.8}$$

in accordance with Lotka's result. However, (6.8) becomes meaningless for the roots with $\Re(s_k) < \sigma$, whereas (6.7) is applicable in all cases.

Theorem 6 can easily be generalized to the case where the *characteristic equation has multiple roots*. The expansion (6.6) (which reduces to the customary expansion into partial fractions whenever $\omega(s)$ is meromorphic) is to be replaced by

$$\omega(s) = \sum_k \left\{ \frac{A_k^{(1)}}{s - s_k} + \frac{A_k^{(2)}}{(s - s_k)^2} + \cdots + \frac{A_k^{(m_k)}}{(s - s_k)^{m_k}} \right\}, \tag{6.9}$$

where m_k is the multiplicity of the root s_k. This leads us formally to an expansion

$$u(t) = \sum_k e^{s_k t} \left\{ A_k^{(1)} + A_k^{(2)} \frac{t}{1!} + \cdots + A_k^{(m_k)} \frac{t^{m_k - 1}}{(m_k - 1)!} \right\}, \tag{6.10}$$

which now replaces (6.2). Generalizing Theorem 6 it is easy to formulate some simple conditions under which (6.11) will really represent a solution of (1.1). Other conditions which ensure that (6.9) is the transform of (6.10) are known from the general theory of Laplace transforms; such conditions usually use only function-theoretical properties of (6.9) and are applicable in particular when $\omega(s)$ is meromorphic. We mention in particular a theorem of Churchill [17] which can be used for our purposes.

7. **On the practical computation of the solution.** There are at hand two main methods for the practical computation of the solution of (1.1). One of them

has been developed by Lotka and consists of an approximate computation of a few coefficients in the series (6.2). The other method uses an expansion

$$u(t) = \sum_{n=0}^{\infty} u_n(t), \tag{7.1}$$

where $u_n(t)$ represents the contribution of the nth "generation" and is defined by x

$$u_0(t) = g(t), \qquad u_{n+1}(t) = \int_0^t u_n(t-x)f(x)\,dx. \tag{7.2}$$

Now the Laplace transform of $u_{n+1}(t)$ is $\gamma(s)\varphi^n(s)$, and hence (7.2) corresponds to the expansion

$$\omega(s) = \frac{\gamma(s)}{1-\varphi(s)} = \gamma(s)\sum_{n=0}^{\infty}\varphi^n(s). \tag{7.3}$$

In practice the functions $g(t)$ and $f(t)$ are usually not known exactly. Frequently their values are obtained from some statistical material, so that only their integrals over some time units, e.g. years, are actually known or, in other words, only the values

$$f_n = \frac{1}{\delta}\int_{n\delta}^{(n+1)\delta} f(t)\,dt, \qquad g_n = \frac{1}{\delta}\int_{n\delta}^{(n+1)\delta} g(t)\,dt, \tag{7.4}$$

are given, where $\delta > 0$ is a given constant. Ordinarily in such cases some theoretical forms (e.g Pearson curves) are fitted to the empirical data and equation (1.1) is solved with these theoretical functions. Now such a procedure is sometimes not only very troublesome, but also somewhat arbitrary. Consider for example the limit of $u(t)$ as $t \to \infty$; this asymptotic value is the main point of interest of the theory and all practical computations. However, as has been shown above, this limit depends only on the moments of the first two orders of $f(t)$ and $g(t)$, and, unless the fitting is done by the method of moments, the resulting value will depend on the special procedure of fitting. Accordingly it will sometimes happen that it is of advantage to use the empirical material as it is, and this can, at least in principle, always be done.

If only the values (7.4) are used it is natural to consider $f(t)$ and $g(t)$ as step-functions defined by

$$\left.\begin{aligned} f(t) &= f_n, \\ g(t) &= g_n, \end{aligned}\right\} \qquad \text{for} \quad n\delta \leq t < (n+1)\delta. \tag{7.5}$$

In practice only a finite number among the f_n and g_n will be different from zero: accordingly the Laplace transforms $\gamma(s)$ and $\varphi(s)$ reduce to trigonometrical polynomials, so that the analytic study of $\omega(s) = \dfrac{\gamma(s)}{1-\varphi(s)}$ becomes particularly simple. Lotka's method can be applied directly in this case.

For a convenient computation of (7.1) it is better to return to the more general equation (1.3), instead of (1.1). The summatory functions $F(t)$ and $G(t)$ should not be defined by (1.2) in this case, but simply by

$$F(t) = \sum_{n=0}^{\left[\frac{t}{\delta}\right]} f_n, \qquad G(t) = \sum_{n=0}^{\left[\frac{t}{\delta}\right]} g_n. \tag{7.6}$$

It is readily seen that the solution $U(t)$ of (1.3) can be written in the form $U(t) = \sum_{n=0}^{\infty} U_n(t)$, where

$$U_0(t) = G(t), \qquad U_{n+1}(t) = \int_0^t U_n(t-x)\,dF(x);$$

in our case $U_n(t)$ will again be a step-function with jumps at the points $k\delta$, the corresponding saltus being

$$u_0^{(k)} = g_k, \qquad u_{n+1}^{(k)} = \sum_{r=0}^{k} u_n^{(k-r)} f_r.$$

Thus we arrive at exactly the same result as would have been obtained if the integrals (7.2) had been computed, starting from (7.4), by the ordinary methods for numerical integration of tabulated functions. It is of interest to note that this method of approximate evaluation of the integrals (7.2) leads to the *exact values of the renewal function* of a population where all changes occur in a discontinuous way at the end of time intervals of length δ in such a way that each change equals the mean value of the changes of the given population over the corresponding time interval.

REFERENCES

I. Papers on the integral equation of renewal theory

Note: Lotka's paper [8] contains a list of 74 papers on the subject published before 1939 The following list is to bring Lotka's list up to June 1941; however no claims to completeness are made

[1] A. W Brown, A note on the use of a Pearson type III function in renewal theory," *Annals of Math. Stat.* Vol. 11 (1940), pp. 448–453.

[2] H. Hadwiger, "Zur Frage des Beharrungszustandes bei kontinuierlich sich erneuernden Gesamtheiten," *Archiv f. mathem. Wirtschafts- und Sozialforschung*, Vol. 5 (1939), pp 32–34.

[3] H. Hadwiger, "Über die Integralgleichung der Bevölkerungstheorie," *Mitteilungen Verein schweizer Versicherungsmathematiker (Bull. Assoc Actuaires suisses)*, Vol. 38 (1939), pp. 1–14.

[4] H. Hadwiger, "Eine analytische Reproduktionsfunktion für biologische Gesamtheiten," *Skand. Aktuarietidskrift* (1940), pp 101–113.

[5] H. Hadwiger, "Natürliche Ausscheidefunktionen für Gesamtheiten und die Lösung der Erneuerungsgleichung," *Mitteilungen Verein. Schweiz. Versich.-Math.*, Vol. 40 (1940), pp 31–39.

[6] H. HADWIGER and W RUCHTI, "Über eine spezielle Klasse analytischer Geburtenfunktionen," *Metron*, Vol. 13 (1939), No. 4, pp 17–26.

[7] A LINDER, "Die Vermehrungsrate der stabilen Bevolkerung," *Archiv f mathem. Wirtschafts- und Sozialforschung*, Vol 4 (1938), pp 136–156

[8] A LOTKA, "A contribution to the theory of self-renewing aggregates, with special reference to industrial replacement," *Annals of Math. Stat*, Vol 10 (1939), pp 1–25

[9] A LOTKA, "On an integral equation in population analysis," *Annals of Math Stat.*, Vol. 10 (1939), pp 144–161

[10] A LOTKA, "Théorie analytique des associations biologiques II," *Actualités Scientifiques No 780*, Paris, 1939.

[11] A LOTKA, "The theory of industrial replacement," *Skand Aktuarietidskrift* (1940), pp 1–14

[12] A LOTKA, "Sur une équation intégrale de l'analyse démographique et industrielle," *Mitt. Verein Schweiz Versich -Math.*, Vol 40 (1940), pp 1–16

[13] H. MUNZNER, "Die Erneuerung von Gesamtheiten," *Archiv f math. Wirtschafts- u. Sozialforschung*, Vol 4 (1938)

[14] G A D. PREINREICH, "The theory of industrial replacement," *Skand Aktuarietidskrift* (1939), pp. 1–19

[15] E. C. RHODES, "Population mathematics, I, II, III," *Roy Stat. Soc Jour*, Vol 103 (1940), pp 61–89, 218–245, 362–387

[16] H RICHTER, "Die Konvergenz der Erneuerungsfunktion," *Blatter f Versicherungsmathematik*, Vol. 5 (1940), pp 21–35.

[16a] H HADWIGER, "Über eine Funktionalgleichung der Bevölkerungstheorie und eine spezielle Klasse analytischer Lösungen," *Bl. f. Versicherungsmathematik*, Vol. 5 (1941), pp 181–188

[16b] G. A. D. PRIENREICH, "The present status of renewal theory," Waverly Press, Baltimore (1940)

II Other papers quoted

[17] R. V. CHURCHILL, "The inversion of the Laplace transformation by a direct expansion in series and its application to boundary-value problems," *Math Zeits*, Vol. 42 (1937), pp 567–579.

[18] G. DOETSCH, *Theorie und Anwendung der Laplace Transformation* J. Springer, Berlin, 1937

[19] W FELLER, "Completely monotone functions and sequences," *Duke Math Jour*, Vol 5 (1939), pp. 661–674

[20] A HAAR, "Über asymptotische Entwicklungen von Funktionen," *Math Ann*, Vol 96 (1927), pp 69–107

[21] R. E. A. C. PALEY and N WIENER, "Notes on the theory and application of Fourier transforms, VII On the Volterra equation," *Amer. Math. Soc. Trans.*, Vol. 35 (1933), pp 785–791

ON THE JOINT DISTRIBUTION OF THE MEDIANS IN SAMPLES FROM A MULTIVARIATE POPULATION

By A. M. Mood

University of Texas

It is well known [1] that in the case of a population having a single variate distributed according to a density function satisfying certain general conditions, the median of a sample is asymptotically normally distributed about the population median as a mean. It is the purpose of this paper to extend this result to populations involving more than one variate. Besides the theoretical interest of such a result, there may be some practical value in it when one is dealing with samples from a population for which the median is a more efficient statistic than the mean, as, for example, when the population variance is not finite.

The complexity of the exact distribution of the sample median increases rapidly with the number of variates which describe the population; it is almost impossible to write out completely the distribution for the general case of k variates For this reason the author has chosen to give first a detailed presentation for the case of two variates, then use a condensed notation to establish the general result. This is a circuitous route, but it seems to be the only feasible one A condensed notation is necessary for the general case, but presented alone it would be well-nigh incomprehensible.

1. Distribution of the median in two dimensions. An extension of A. T. Craig's [2] geometrical argument will be used to obtain the exact distribution of the sample median. Let us consider two variates x_1 and x_2 with density function $f(x_1, x_2)$ which shall satisfy the following conditions:

1. $f(x_1, x_2) \geq 0$
2. $$\int_{-\infty}^{\infty} f\left(x_1, \frac{1}{N}\right) dx_1 = \int_{-\infty}^{\infty} f(x_1, 0)\, dx_1 + O\left(\frac{1}{N}\right)$$
   $$\int_{-\infty}^{\infty} f\left(\frac{1}{N}, x_2\right) dx_2 = \int_{-\infty}^{\infty} f(0, x_2)\; dx_2 + O\left(\frac{1}{N}\right)$$
3. $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1\, dx_2 = 1$$
4. Each of the equations
   $$\int_{-\infty}^{\xi}\int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 dx_1 = \tfrac{1}{2}$$
   $$\int_{-\infty}^{\xi}\int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1 dx_2 = \tfrac{1}{2}$$
   has a unique real root.

If ξ_1 and ξ_2 are the respective roots of the two equations of this last condition then the point (ξ_1, ξ_2) is defined to be the population median. It will be assumed in what follows that the coordinate system has been so chosen that $\xi_1 = 0 = \xi_2$.

Let a sample of $2n + 1$ elements $(x_{1\alpha}, x_{2\alpha})(\alpha = 1, 2, \cdots, 2n + 1)$ be drawn from this population. The sample median $(\tilde{x}_1, \tilde{x}_2)$ will be defined as an element (not necessarily in the sample) whose x_1 coordinate is the middle, with respect to magnitude, number of the set of numbers $x_{1\alpha}$, and whose x_2 coordinate is the middle number of the set of numbers $x_{2\alpha}$. Now let us compute the probability that the sample median will lie in the rectangle

$$\tilde{x}_i - \tfrac{1}{2}\, d\tilde{x}_i < x_i < \tilde{x}_i + \tfrac{1}{2}\, d\tilde{x}_i \qquad i = 1, 2.$$

This rectangle will be denoted by R''. The remainder of the plane will be divided into eight other regions $R_1, \cdots, R_4'$ as indicated by the dotted lines in Figure 1. The probability that an element will fall in the region $R_i^{(j)}$ will be denoted by

$$p_i^{(j)} = \int\int_{R_i^{(j)}} f(x_1, x_2)\, dx_1 dx_2 .$$

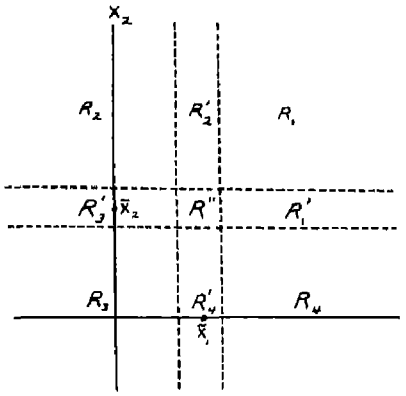


FIG. 1

Neglecting terms involving differentials of higher order we have

$$\begin{aligned}
p_1 &= \int_{\tilde{x}_1}^{\infty}\int_{\tilde{x}_2}^{\infty} f(x_1, x_2)\, dx_2 dx_1 \\
p_2 &= \int_{-\infty}^{\tilde{x}_1}\int_{\tilde{x}_2}^{\infty} f(x_1, x_2)\, dx_2\, dx_1 \\
&\vdots \\
p' &= \int_{\tilde{x}_1}^{\infty} f(x_1, \tilde{x}_2)\, dx_1\, d\tilde{x}_2 \\
&\vdots \\
p'' &= f(\tilde{x}_1, \tilde{x}_2)\, d\tilde{x}_1 d\tilde{x}_2 .
\end{aligned} \tag{1}$$

We shall consider now that the sample is drawn from a multinomial population with probabilities $p_1, \cdots, p''$ and pick out those terms which give rise to a sample median in R''. If the median is an element of the sample, then that element must fall in R'' and the other elements must fall in the regions R_1, R_2, R_3, and R_4 in such a manner that

$$n_1 + n_2 = n_3 + n_4 = n$$

$$n_1 + n_4 = n_2 + n_3 = n$$

or so that

$$n_1 = n_3 \text{ and } n_2 = n_4 \tag{2}$$

where n_i is the number of elements in R_i. The probability that this occurs is

$$\sum_{n_1+n_2=n} \frac{(2n+1)!}{n_1!^2\, n_2!^2}\, p''p_1^{n_1}p_2^{n_2}p_3^{n_1}p_4^{n_2} \tag{3}$$

Now suppose the median is determined by two different elements of the sample, for example one in R_1' and one in R_2', then there must be n_1 elements in R_1, $n_1 + 1$ elements in R_3, and n_2 elements in each of R_2 and R_4 with

$$n_1 + n_2 = n - 1. \tag{4}$$

The probability in this case is

$$p_1'p_2' \sum_{n_1+n_2=n-1} \frac{(2n+1)!}{n_1!\,(n_1+1)!\,n_2!^2}\, p_1^{n_1}p_2^{n_2}p_3^{n_1+1}p_4^{n_2}. \tag{5}$$

Continuing in this manner we obtain the distribution of the median, and letting $D(\tilde{x}_1, \tilde{x}_2)$ represent the density function giving this distribution we have

$$\begin{aligned} D(\tilde{x}_1, \tilde{x}_2)\, d\tilde{x}_1\, d\tilde{x}_2 = {} & p'' \Sigma \frac{(2n+1)!}{n_1!^2\, n_2!^2}\, (p_1p_3)^{n_1}(p_2p_4)^{n_2} \\ & + (p_3p_1'p_2' + p_1p_3'p_4') \Sigma \frac{(2n+1)!}{n_1!\,(n_1+1)!\,n_2!^2}\, (p_1p_3)^{n_1}(p_2p_4)^{n_2} \\ & + (p_2p_1'p_4' + p_4p_2'p_3') \Sigma \frac{(2n+1)!}{n_1!^2\, n_2!\,(n_2+1)!}\, (p_1p_3)^{n_1}(p_2p_4)^{n_2}. \end{aligned} \tag{6}$$

2. Asymptotic distribution of the median in two dimensions. As a simple notation

$$A = B(1 \dot{+} O(1/\sqrt{n}))$$

will be abbreviated to read

$$A =\cdot B, \tag{7}$$

the dot after the equality sign indicating the omission of the factor $1 + O(1/\sqrt{n})$.

As is customary, the second term of this factor represents any function such that

$$\lim_{N\to\infty} NO(1/N) = L < \infty.$$

In order to get an approximation to (6) for large n we shall use the normal approximation for the multinomial distribution and compute the sums (these cannot be put in finite form) by integration. We use then the well-known result

$$\frac{m!}{\prod_1^r m_i!}\prod_1^r p_i^{m_i} = \cdot\,[A/(2\pi)^{r-1}]^{\frac{1}{2}} \exp\left(-\tfrac{1}{2}\sum_1^{r-1} A_{ij}z_iz_j\right)\prod_1^{r-1} dz_i, \tag{8}$$

where

$$z_i = (m_i - mp_i)/\sqrt{m}, \qquad i = 1, 2, \cdots, r-1, \tag{9}$$

$$A_{ii} = \frac{1}{p_i} + \frac{1}{p_r}, \qquad A_{ij} = \frac{1}{p_r}. \tag{10}$$

Returning to (6) it is to be noted that the fraction immediately following Σ in the first sum has one more factor in the denominator than the corresponding fractions in the other sums This first sum may therefore be neglected in the asymptotic form as it is of order $1/n$ in comparison with the others We consider now the second sum in (6) and let it be represented by the letter S

$$S = 2n(2n+1)p_1'p_2' \sum_{n_1+n_2=n-1} \frac{(2n-1)}{n_1!\,(n_1+1)!\,n_2!^2} p_1^{n_1}p_2^{n_2}p_3^{n_1+1}p_4^{n_2}. \tag{11}$$

Employing (8) and omitting certain terms of order $1/n$ we have

$$S = \cdot\, 4n^2p_1'p_2' \sum [A/(2\pi)^3]^{\frac{1}{2}} \exp\left(-\tfrac{1}{2}\sum_1^3 A_{ij}z_iz_j\right) dz_1\,dz_2\,dz_3, \tag{12}$$

in which the A_{ij} are defined by (10) with $r = 4$, and

$$z_i = (n_i - 2np_i)/\sqrt{2n}, \qquad i = 1, 2, 3. \tag{13}$$

In view of the relations (2) between the n_i we have

$$\begin{aligned} z_2 &= \sqrt{2n}\,(\tfrac{1}{2} - p_1 - p_2) - z_1 = u_1 - z_1 \\ z_3 &= \sqrt{2n}\,(p_1 - p_3) - z_1 = u_2 - z_1, \end{aligned} \tag{14}$$

in which relations we have defined the new symbols u_1 and u_2. It will be recalled that in (8) the factors dz_i correspond to factors $1/\sqrt{m}$, we therefore let dz_2 and dz_3 in (12) cancel a factor $2n$ from the coefficient of the exponential, and after substituting (14) in (12) find that

$$\begin{aligned} S = \cdot\, 2np_1'p_2' \Sigma\,[A/(2\pi)^3]^{\frac{1}{2}} \exp\Big\{-\tfrac{1}{2}\Big[z_1^2\Big(\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p_3}+\frac{1}{p_4}\Big) \\ + 2z_1\Big(\frac{u_1+u_2}{p_4} - \frac{u_1}{p_2} + \frac{u_2}{p_3}\Big) + \frac{(u_1+u_2)^2}{p_4} + \frac{u_1^2}{p_2} + \frac{u_2^2}{p_3}\Big]\Big\} dz_1. \end{aligned} \tag{15}$$

The summation can now be performed to within terms of order $1/\sqrt{n}$ by integration with respect to z_1 between the limits $-\infty$ and $+\infty$; this gives us

$$S = \cdot \frac{2np_1'p_2'}{2\pi} A^{\frac{1}{2}} \Big/ \left(\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} + \frac{1}{p_4}\right)^{\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\left[\frac{(u_1+u_2)^2}{p_4} + \frac{u_1^2}{p_2} + \frac{u_2^2}{p_3} - \left(\frac{u_1+u_2}{p_4} - \frac{u_1}{p_2} + \frac{u_2}{p_3}\right)^2 \Big/ \left(\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} + \frac{1}{p_4}\right)\right]\right\}. \tag{16}$$

At this point some new symbols are required. We let q_i and q_i' represent the results of replacing $\tilde{x}_1$ and $\tilde{x}_2$ by zero in the integrals of the relations (1)

$$\begin{aligned} q_1 &= \int_0^\infty \int_0^\infty f(x_1, x_2)\, dx_1\, dx_2 & q_1' &= \int_0^\infty f(x_1, 0)\, dx_1 \\ q_2 &= \int_{-\infty}^0 \int_0^\infty f(x_1, x_2)\, dx_1\, dx_2 & q_2' &= \int_0^\infty f(0, x_2)\, dx_2 \\ q_3 &= \int_{-\infty}^0 \int_{-\infty}^0 f(x_1, x_2)\, dx_1\, dx_2 & q_3' &= \int_{-\infty}^0 f(x_1, 0)\, dx_1 \\ q_4 &= \int_0^\infty \int_{-\infty}^0 f(x_1, x_2)\, dx_1\, dx_2 & q_4' &= \int_{-\infty}^0 f(0, x_2)\, dx_2 \end{aligned} \tag{17}$$

then

$$q_1 + q_2 = q_3 + q_4 = q_1 + q_4 = q_2 + q_3 = \tfrac{1}{2} \tag{18}$$

and

$$q_1 = q_3\,, \qquad q_2 = q_4\,. \tag{19}$$

Also we let

$$a_1 = q_2' + q_3'\,, \qquad a_2 = q_1' + q_4'\,, \tag{20}$$

$$y_1 = \sqrt{2n}\, a_1 \tilde{x}_1\,, \qquad y_2 = \sqrt{2n}\, a_2 \tilde{x}_2\,. \tag{21}$$

We have now

$$\begin{aligned} p_i &= \cdot\, q_i\,, & i &= 1, 2, 3, 4, \\ p_i' &= \cdot\, q_i' d\tilde{x}_2\,, & i &= 1, 3, \\ p_i' &= \cdot\, q_i' d\tilde{x}_1\,, & i &= 2, 4. \end{aligned} \tag{22}$$

Also

$$\begin{aligned} u_1 &= \sqrt{2n}\,(\tfrac{1}{2} - p_1 - p_2) \\ &= \sqrt{2n} \int_{-\infty}^\infty \int_0^{\tilde{x}_2} f(x_1, x_2)\, dx_1\, dx_2 \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{2n}\,\tilde{x}_2 \int_{-\infty}^{\infty} f(x_1, \theta\tilde{x}_2)\,dx_1, \qquad 0 \leq \theta \leq 1, \\
&=\cdot \sqrt{2n}\,\tilde{x}_2 \int_{-\infty}^{\infty} f(x_1, 0)\,dx_1 \\
&=\cdot \sqrt{2n}\,a_2\tilde{x}_2 \\
&=\cdot y_2 .
\end{aligned} \tag{23}
$$

Similarly

$$
\begin{aligned}
u_2 &= \sqrt{2n}\,(p_1 - p_3) \\
&=\cdot -(y_1 + y_2).
\end{aligned} \tag{24}
$$

The result of substituting (22), (23) and (24) in (16) with some further simplification using (18) and (19) is

$$
S =\cdot \frac{2nq_1'q_2'}{2\pi\sqrt{q_1q_2}} \exp\left(-\frac{1}{2}\frac{y_1^2 - 4(q_1 - q_2)y_1y_2 + y_2^2}{4q_1q_2}\right) d\tilde{x}_1\,d\tilde{x}_2 . \tag{25}
$$

The other three sums of (6) will give rise to the same expression except that the factors $q_1'q_2'$ will be different; it is clear then that

$$
\begin{aligned}
D(\tilde{x}_1, \tilde{x}_2)\,d\tilde{x}_1\,d\tilde{x}_2 &=\cdot \frac{2n(q_1'q_2' + q_1'q_2' + q_2'q_3' + q_3'q_4')}{2\pi\sqrt{q_1q_2}} \\
&\qquad \times \exp\left(-\frac{1}{2}\frac{y_1^2 - 4(q_1 - q_2)y_1y_2 + y_2^2}{4q_1q_2}\right) d\tilde{x}_1\,d\tilde{x}_2 \\
&=\cdot \frac{2na_1a_2}{2\pi\sqrt{q_1q_2}} \exp\left(-n\frac{a_1^2\tilde{x}_1^2 - 4(q_1 - q_2)a_1a_2\tilde{x}_1\tilde{x}_2 + a_2^2\tilde{x}_2^2}{4q_1q_2}\right) d\tilde{x}_1\,d\tilde{x}_2, \qquad (26) \\
&=\cdot \frac{1}{2\pi\sqrt{q_1q_2}} \exp\left(-\frac{1}{2}\frac{y_1^2 - 4(q_1 - q_2)y_1y_2 + y_2^2}{4q_1q_2}\right) dy_1\,dy_2 . \qquad (27)
\end{aligned}
$$

This is the asymptotic form for the distribution of the median in two dimensions.

3. Distribution of the median in k dimensions. We consider now a population characterized by a density function $f(x_1, \cdots, x_k)$ defined over a euclidean space of k dimensions satisfying conditions like those required of $f(x_1, x_2)$ in section 1, and we assume that the population median is at the origin so that the integral of the density function over any half-space determined by a coordinate hyperplane is $\frac{1}{2}$.

A sample of $2n + 1$ elements will have a median $(\tilde{x}_1, \cdots, \tilde{x}_k)$ each coordinate of which is the middle number of the set of numbers giving the corresponding coordinate of the elements of the sample. To obtain the probability that the sample median lies in the hyperparallopiped $\tilde{x}_\alpha - \frac{1}{2}d\tilde{x}_\alpha < x_\alpha < \tilde{x}_\alpha + \frac{1}{2}\,d\tilde{x}_\alpha$ $(\alpha = 1, 2, \cdots, k)$, we divide the space into 3^k regions by means of hyperplanes

perpendicular to the coordinate axes through the points $\tilde{x}_\alpha \pm \frac{1}{2}\, d\tilde{x}_\alpha$ on the coordinate axes. These regions are illustrated in Figure 2 for the case of three dimensions. The coordinate axes have been omitted in this figure. There will be 2^k primary regions denoted by $R_1, R_2, \cdots, R_{2^k}$ corresponding to the octants of the figure; $k2^{k-1}$ regions with one differential dimension denoted by $R_1', R_2', \cdots, R_{k2^{k-1}}'$ corresponding to the quarter slabs of the figure; $\binom{k}{2}2^{k-1}$ regions with two differential dimensions corresponding to the half strips of the figure, and so forth. Probabilities associated with these regions are defined by

$$p_i^{(j)} = \int_{R_i^{(j)}} f(x_1, \cdots, x_k)\, dx_1 \cdots dx_k .$$

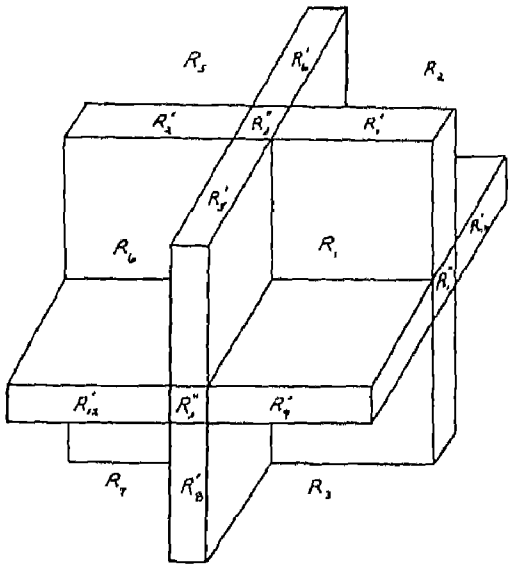


FIG. 2

If the sample median is determined by k different elements of the sample there will be one of these k elements in each of k regions R_i' whose differential dimensions are mutually orthogonal and the other elements of the sample will fall in the regions R_i in such a way that n elements of the sample will lie on either side of any of the k hyperplanes $x_\alpha = \tilde{x}_\alpha$. The probability of this occurrence for a particular choice of k of the regions R_i' is

$$S = \prod_{\alpha=1}^{k} p_{i_\alpha}' \sum \frac{(2n+1)!}{\prod n_i!} \prod_{i=1}^{2^k} p_i^{n_i} \tag{28}$$

in which the 2^k indices n_i are subject to k independent restrictions of the type

$$\Sigma' n_i = n - c_\alpha , \tag{29}$$

where c_α is an integer such that $0 \leq c_\alpha < k$, and the prime on Σ indicates that the sum is to be taken over all n_i on one side of a hyperplane $x_\alpha = \tilde{x}_\alpha$. n_i is the number of elements in R_i and besides the k conditions (29) we have also

$$\sum_1^{2^k} n_i = 2n - k + 1. \tag{30}$$

In order to include all ways in which the median is determined by k different elements of the sample we must add together $2^{k(k-1)}$ sums of the type (28). If the median is determined by less than k elements, say $k - h$ elements, then the fraction $(2n + 1)!/\Pi n_i!$ will have h extra factors in the denominator and hence the sum will be of order $1/n^h$ as compared with that of (28) and may be neglected in obtaining an asymptotic expression

Thus we need only find the limiting form of (28)

$$S = (2n + 1)(2n) \cdots (2n - k + 2) \prod_1^k p'_{i_\alpha} \sum \frac{(2n - k + 1)!}{\prod n_i!} \prod_1^{2^k} p_i^{n_i},$$

which after substituting (8) and neglecting terms of lower order becomes

$$S = \cdot (2n)^k \prod p'_{i_\alpha} \sum (A/(2\pi)^{2^k-1})^{\frac{1}{2}} \exp\left(-\tfrac{1}{2} \sum A_{ij} z_i z_j\right) \prod_1^{2^k-1} dz_i, \tag{31}$$

in which the A_{ij} are defined by (10) with $r = 2^k$ and

$$z_i = (n_i - 2np_i)/\sqrt{2n}, \qquad i = 1, 2, \cdots, 2^k - 1. \tag{32}$$

Now we define

$$u_\alpha = \sqrt{2n}(\tfrac{1}{2} - \Sigma' p_i), \qquad \alpha = 1, 2, \cdots, k, \tag{33}$$

the Σ' having the same significance as in (29). These conditions (29) may now be put in the form

$$z_\alpha = \cdot u_\alpha - L_\alpha(z),$$

in which $L_\alpha(z)$ is a sum of a certain subset of the variables $z_{k+1}, \cdots, z_{2^k-1}$. Care must be taken in labeling the regions R_i in order to be able to solve for $z_1, \cdots, z_k$ in this form. After substituting these relations in (31) we replace $\prod_1^k dz_\alpha$ by $(1/2n)^{k/2}$ and perform the summation to within terms of order $1/\sqrt{n}$ by integrating the remaining z_i from $-\infty$ to $+\infty$; the result is

$$S = \cdot (2n/2\pi)^{k/2} \prod_{\alpha=1}^k p'_{i_\alpha} \sqrt{B} \exp\left(-\tfrac{1}{2} \sum_1^k B_{\alpha\beta} u_\alpha u_\beta\right), \tag{34}$$

in which the $B_{\alpha\beta}$ are functions of the p_i, and $B = |B_{\alpha\beta}|$. As in (17) and (20)

we define

$$q_i = \int_{\bar{R}_i} f(x_1, \cdots, x_k)\Pi\, dx_\alpha$$

$$q_i' = \int_{\bar{R}_i'} f(x_1, \cdots, x_k)\Pi'\, dx_\alpha \tag{35}$$

$$a_\alpha = \int_{x_\alpha=0} f(x_1, \cdots, x_k)\Pi'\, dx_\alpha = \Sigma' q_i',$$

in which $\bar{R}_i$ is the set of regions bounded by the coordinate hyperplanes R_i' are regions into which the coordinate hyperplanes are divided by the remaining coordinate hyperplanes. Π' indicates that one of the differentials is omitted and the variate corresponding to that differential is put equal to zero in $f(x_1, \cdots, x_k)$; Σ' indicates the sum over all q' determined by regions lying in the hyperplane $x_\alpha = 0$. It is clear that

$$p_i = \cdot q_i$$

$$\prod_\alpha p_{i_\alpha}' = \cdot \prod_\alpha q_{i_\alpha}'\, d\tilde{x}_\alpha \tag{36}$$

$$u_\alpha = \cdot \sqrt{2n} \sum_{\beta=1}^{k} \delta_{\alpha\beta} a_\beta \tilde{x}_\beta = \sum \delta_{\alpha\beta} y_\beta,$$

where

$$\delta_{\alpha\beta} = \pm 1 \text{ or } 0, \qquad \text{and} \qquad \gamma_\beta = \sqrt{2n} a_\beta \tilde{x}_\beta .$$

Making these substitutions in (34) we have

$$S = \cdot (2n/2\pi)^{k/2} \prod_1^k q_{i_\alpha}' \sqrt{C} \exp\left(-n \sum_1^k C_{\alpha\beta} a_\alpha a_\beta \tilde{x}_\alpha \tilde{x}_\beta\right) \prod d\tilde{x}_\alpha, \tag{37}$$

and adding together all possible sums of the type (28) we have the asymptotic form of the distribution of the sample median

$$D(\tilde{x}_1, \cdots, \tilde{x}_k) \prod d\tilde{x}_\alpha \tag{38}$$

$$= \cdot (2n/2\pi)^{k/2} \prod_1^k a_\alpha \sqrt{C} \exp\left(-n \sum_1^k C_{\alpha\beta} a_\alpha a_\beta \tilde{x}_\alpha \tilde{x}_\beta\right) \prod d\tilde{x}_\alpha$$

$$= \cdot (1/2\pi)^{k/2} \sqrt{C} \exp\left(-\tfrac{1}{2} \sum C_{\alpha\beta} y_\alpha y_\beta\right) \prod dy_\alpha, \tag{39}$$

in which the $C_{\alpha\beta}$ are functions of the q_i.

4. The case of three dimensions. The computation of the coefficients $C_{\alpha\beta}$ of (39) requires the evaluation of a determinant of order $2^k - k$ for each one of them. This work was quite laborious even for $k = 3$ and the author made no attempt to find their explicit expression for larger values of k.

If we let a subscript $+$ indicate integration of the density function $f(x_1, x_2, x_3)$ from 0 to ∞, and a subscript—indicate integration from $-\infty$ to 0,

as for example,

$$f_{++-} = \int_0^\infty \int_0^\infty \int_{-\infty}^0 f(x_1, x_2, x_3)\, dx_3\, dx_2\, dx_1,$$

then the q_i of (35) will be defined as follows

$$\begin{aligned} q_1 &= f_{+++} & q_5 &= f_{-++} \\ q_2 &= f_{++-} & q_6 &= f_{-+-} \\ q_3 &= f_{+-+} & q_7 &= f_{--+} \\ q_4 &= f_{+--} & q_8 &= f_{---} \end{aligned} \tag{40}$$

The coefficients $C_{\alpha\beta}$ may be written

$$\begin{aligned} DC_{11} &= 2(q_1 + q_5)(q_2 + q_6) \\ DC_{22} &= 2(q_1 + q_3)(q_2 + q_4) \\ DC_{33} &= 2(q_1 + q_2)(q_3 + q_4) \\ DC_{12} &= q_3q_5 + q_4q_6 - q_1q_7 - q_2q_8 \\ DC_{13} &= q_2q_5 + q_4q_7 - q_1q_6 - q_3q_8 \\ DC_{23} &= q_2q_3 + q_6q_7 - q_1q_4 - q_5q_8\,, \end{aligned} \tag{41}$$

where

$$\begin{aligned} D = q_1q_2q_3q_4\left(\frac{1}{q_1} + \frac{1}{q_2} + \frac{1}{q_3} + \frac{1}{q_4}\right) + q_5q_6q_7q_8\left(\frac{1}{q_5} + \frac{1}{q_6} + \frac{1}{q_7} + \frac{1}{q_8}\right) \\ + 2(q_5 + q_6)(q_7 + q_8)(q_1q_2 + q_3q_4) \\ + 2(q_5 + q_7)(q_6 + q_8)(q_1q_3 + q_2q_4) \\ + 2(q_5 + q_8)(q_6 + q_7)(q_1q_4 + q_2q_3) \\ + 8(q_1q_4q_6q_7 + q_2q_3q_5q_8) \end{aligned} \tag{42}$$

(41) and (42) can of course be put in different forms by using the four relations between the q_i. The a_α of (38) are defined in (35); for $k = 3$ they are

$$\begin{aligned} a_1 &= \int_{-\infty}^\infty \int_{-\infty}^\infty f(0, x_2, x_3)\, dx_2\, dx_3 \\ a_2 &= \int_{-\infty}^\infty \int_{-\infty}^\infty f(x_1, 0, x_3)\, dx_1\, dx_3 \\ a_3 &= \int_{-\infty}^\infty \int_{-\infty}^\infty f(x_1, x_2, 0)\, dx_1\, dx_2\,. \end{aligned} \tag{43}$$

5. The normal distribution in two dimensions. If the density function of the second section of the paper is normal

$$(44)\quad f(x_1, x_2) = 1/(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}) \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2} - 2\rho\frac{x_1x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right],$$

we find that the parameters of (26) are

$$(45)\quad \begin{aligned} q_1 &= \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho, & q_2 &= \frac{1}{4} - \frac{1}{2\pi}\sin^{-1}\rho, \\ a_1 &= \frac{1}{\sqrt{2\pi}\,\sigma_1}, & a_2 &= \frac{1}{\sqrt{2\pi}\,\sigma_2}. \end{aligned}$$

These give an interesting result—the correlation coefficient of the asymptotic distribution of the sample medians is

$$(46)\quad \rho_m = \frac{2}{\pi}\sin^{-1}\rho$$

hence

$$(47)\quad |\rho_m| \leq |\rho|$$

the equality sign holding only when $\rho = 0$ or ± 1.

REFERENCES

[1] S. S. Wilks, *Statistical Inference*, Ann Arbor, Edwards Bros., 1937.

[2] A. T. Craig, "On the distributions of certain statistics," *Amer. Jour. Math.*, Vol. 54 (1932), pp. 353-366.

SAMPLES FROM TWO BIVARIATE NORMAL POPULATIONS[1]

By Chung Tsi Hsu

Columbia University

1. **Introduction.** In multivariate analysis involving p variates, or in analysis of variance of m samples from univariate populations, we are often interested in the hypothesis of the equality of variances, viz., that

$$\sigma_1 = \sigma_2 = \cdots = \sigma_p, \qquad \text{in the case of } p \text{ variates;}$$

or

$$\sigma_1 = \sigma_2 = \cdots = \sigma_m, \qquad \text{in the case of } m \text{ samples.}$$

As a matter of fact, it seldom occurs that these hypotheses are true, but the ratio between the variances might be known.

Hotelling [5] has suggested that if

$$\sigma_1^2/k_1 = \sigma_2^2/k_2 = \cdots = \sigma_m^2/k_m = \sigma^2,$$

where the k's are known constants, we can apply the transformation

$$\begin{aligned} x_1' &= w_1 x_1, \\ x_2' &= w_2 x_2, \\ &\cdots\cdots\cdots \\ x_m' &= w_m x_m, \end{aligned}$$

where

$$w\sqrt{k_1} = w_2\sqrt{k_2} = \cdots = w_m\sqrt{k_m} = 1,$$

so that after transformation the variances become equal, i.e.,

$$\sigma_1' = \sigma_2' = \cdots = \sigma_m',$$

and the required analysis can be carried out. This method is similarly applicable in the multivariate case.

In a previous paper [7], I developed a series of hypotheses concerning samples from a bivariate normal population under the assumption that

$$\sigma_1 = \sigma_2.$$

In case $\sigma_1^2/k_1 = \sigma_2^2/k_2$, where k_1 and k_2 are two distinct known constants, similar results may be obtained by the use of the transformation $x_1' = w_1 x_1$; $x_2' = w_2 x_2$; where $w_1\sqrt{k_1} = w_2\sqrt{k_2} = 1$.

[1] Presented to the American Mathematical Society at Washington, D. C., May 3, 1941.

In multivariate analysis, the hypotheses usually of interest concerning correlation coefficients may be classified in two categories, viz.,

(i) that the correlation coefficient is equal to a specified value, e.g., in simple correlation $\rho_{12} = \rho_0$, in partial correlation, $\rho_{12.3} = \rho_0$, in multiple correlation, $\rho_{1.23} = \rho_0$, or in correlation between two sets of variates [4][2], $Q = Q_0$; of special interest is the hypothesis of the vanishing of such correlation coefficients.

(ii) that two given correlation coefficients are equal, e.g., (1) correlation coefficients ρ_1 and ρ_2 in the correlation matrix of a multivariate distribution are equal (Hotelling [6]), or (2) the correlation coefficients ρ_{12} and ρ'_{12} in two bivariate populations are equal.

R. A. Fisher in his earlier paper [3] introduced the transformation $z = \frac{1}{2}\log\frac{1+r}{1-r}$ which provides a very satisfactory, though approximate, method for the comparison of two correlation coefficients. Brander [1] treated the same problem by the method of the likelihood ratio criterion.

The present paper is an attempt to obtain different criteria by the likelihood ratio method (Neyman and Pearson [9], [10], [11]) for testing, by means of samples, the equality of correlation coefficients in two bivariate normal populations under the following sets of conditions: (1) $\sigma_1 = \sigma_2$ and $\sigma'_1 = \sigma'_2$; (2) $\sigma_1 = \sigma_2$, $\xi_1 = \xi_2$ and $\sigma'_1 = \sigma'_2$, $\xi'_1 = \xi'_2$. The results may be extended to the cases (3) $\sigma_1^2/k_1 = \sigma_2^2/k_2$ and $\sigma_1'^2/k_1' = \sigma_2'^2/k_2'$; (4) $\sigma_1^2/k_1 = \sigma_2^2/k_2$, $\xi_1^2/k_1 = \xi_2^2/k_2$ and $\sigma_1'^2/k_1' = \sigma_2'^2/k_2'$, $\xi_1'^2/k_1' = \xi_2'^2/k_2'$, where the k's are known constants.

2. **The hypotheses.** Two samples, each being of two variates (x_1, x_2) and (x'_1, x'_2), of size N and N', are supposed to be drawn at random, respectively, from two independent normal bivariate populations, with the following distributions:

$$(1)\qquad \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\xi_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\xi_1}{\sigma_1}\right)\left(\frac{x_2-\xi_2}{\sigma_2}\right)+\left(\frac{x_2-\xi_2}{\sigma_2}\right)^2\right]\right\},$$

$$(2)\qquad \frac{1}{2\pi\sigma'_1\sigma'_2\sqrt{1-\rho'^2}}\exp\left\{-\frac{1}{2(1-\rho'^2)}\left[\left(\frac{x'_1-\xi'_1}{\sigma'_1}\right)^2 - 2\rho'\left(\frac{x'_1-\xi'_1}{\sigma'_1}\right)\left(\frac{x'_2-\xi'_2}{\sigma'_2}\right)+\left(\frac{x'_2-\xi'_2}{\sigma'_2}\right)^2\right]\right\},$$

where $\xi_1, \xi_2, \sigma_1, \sigma_2, \rho; \xi'_1, \xi'_2, \sigma'_1, \sigma'_2, \rho'$ are the unknown parameters of the populations.

The hypotheses to be considered in the present paper are:

H_1: Assuming $\sigma_1 = \sigma_2$ and $\sigma'_1 = \sigma'_2$, to test $\rho = \rho'$.

H_2: Assuming $\sigma_1 = \sigma_2$, $\xi_1 = \xi_2$, and $\sigma'_1 = \sigma'_2$, $\xi'_1 = \xi'_2$, to test $\rho = \rho'$.

[2] See bibliography at the end of the paper.

The derivation and the distribution of the criteria for testing these hypotheses may be simplified by the following simultaneous transformations:

$$X = \frac{1}{\sqrt{2}}(x_1 - x_2) \qquad Y = \frac{1}{\sqrt{2}}(x_1 + x_2) \tag{3}$$

$$X' = \frac{1}{\sqrt{2}}(x_1' - x_2') \qquad Y' = \frac{1}{\sqrt{2}}(x_1' + x_2') \tag{4}$$

The corresponding normal bivariate distributions in the transformed variables (X, Y) and (X', Y') are obtained, viz.

$$\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \exp\left\{-\frac{1}{2(1-\rho_{XY}^2)}\left[\left(\frac{X-\xi}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{X-\xi}{\sigma_X}\right)\left(\frac{Y-\eta}{\sigma_Y}\right) + \left(\frac{Y-\eta}{\sigma_Y}\right)^2\right]\right\} dX\,dY, \tag{5}$$

$$\frac{1}{2\pi\sigma_X'\sigma_Y'\sqrt{1-\rho_{XY}'^2}} \exp\left\{-\frac{1}{2(1-\rho_{XY}'^2)}\left[\left(\frac{X'-\xi'}{\sigma_X'}\right)^2 - 2\rho_{XY}'\left(\frac{X'-\xi'}{\sigma_X'}\right)\left(\frac{Y'-\eta'}{\sigma_Y'}\right) + \left(\frac{Y'-\eta'}{\sigma_Y'}\right)^2\right]\right\} dX'\,dY'. \tag{6}$$

The conditions corresponding to

$$\sigma_1 = \sigma_2 \quad \text{and} \quad \sigma_1' = \sigma_2', \tag{7}$$

are that

$$\rho_{XY} = 0 \quad \text{and} \quad \rho_{XY}' = 0. \tag{8}$$

Also, for a given ρ and ρ', we have from (7)

$$\sigma_Y^2 = \gamma\sigma_X^2 \quad \text{and} \quad \sigma_Y'^2 = \gamma'\sigma_X'^2, \tag{9}$$

where

$$\gamma = \frac{1+\rho}{1-\rho} \qquad \text{and} \qquad \gamma' = \frac{1+\rho'}{1-\rho'}. \tag{10}$$

Following the notation of (9) and (10), the hypotheses H_1' and H_2' corresponding to H_1 and H_2 are:

H_1' : Assuming $\rho_{XY} = 0$, and $\rho_{XY}' = 0$, to test $\gamma = \gamma'$.

H_2' : Assuming $\rho_{XY} = 0$, $\xi = 0$, and $\rho_{XY}' = 0, \xi' = 0$, to test $\gamma = \gamma'$.

3. The derivation of the criteria. Let $(x_{1i}, x_{2i})(x_{1j}', x_{2j}')$ be the measurements of the characters on the ith and jth individuals in the two samples from their respective populations After transformation, the corresponding measurements become (X_i, Y_i) and (X_j', Y_j'). Let $p(E)$ denote the joint elementary proba-

bility law of the N and N' observations, $E = (X_1, \cdots, X_N, Y_1, \cdots, Y_N; X'_1, \cdots, X'_{N'}, Y', \cdots, Y'_{N'})$.

Following Neyman and Pearson, we shall use Ω to designate the class of admissible populations under conditions which can be assumed to be satisfied in any case; and ω to designate a subclass of Ω under conditions which are satisfied only if the hypothesis to be tested is true.

Thus for H', Ω specifies for $\rho_{XY} = \rho'_{XY} = 0$, any real values of ξ, η, ξ', η' and any positive values of σ_X, σ_Y, σ'_X, σ'_Y; ω specifies $\rho_{XY} = \rho'_{XY} = 0$, any real values of ξ, η, ξ', η' and any positive values of σ_Y and γ which are defined by (9). While for H', Ω specifies $\rho_{XY} = \rho'_{XY} = 0$, $\xi = \xi' = 0$, any real values of η and η' and any positive values of σ_X, σ_Y, σ'_X, σ'_Y; ω specifies $\rho_{XY} = \rho'_{XY} = 0$, $\xi = \xi' = 0$, any real values of η and η', and any positive values of σ_Y and γ which are defined by (9).

For our hypothesis H'_1, the values of the parameters required to make $p(\Omega)$ a maximum are:

$$\hat{\xi} = \bar{X}, \qquad \hat{\eta} = \bar{Y}, \qquad \hat{\sigma}_X = s_X, \qquad \hat{\sigma}_Y = s_Y$$
$$\hat{\xi}' = \bar{X}', \qquad \hat{\eta}' = \bar{Y}', \qquad \hat{\sigma}'_X = s'_X, \qquad \hat{\sigma}'_Y = s'_Y.$$

Thus $p(\Omega \max) = \left(\frac{1}{2\pi}\right)^{N+N'} \frac{1}{s_X^N s_Y^N s_X'^{N'} s_Y'^{N'}} e^{-N-N'}$

To obtain $p(\omega \max)$, let us define, according to the notation in the writer's previous paper [7],

$$R_1 = \frac{2Y s_1 s_2}{s_1^2 + s_2^2} \qquad \text{and} \qquad R'_1 = \frac{2Y' s'_1 s'_2}{s_1'^2 + s_2'^2}$$

and

$$u = \frac{s_Y^2}{s_X^2} = \frac{1 + R_1}{1 - R_1} \qquad u' = \frac{s_Y'^2}{s_X'^2} = \frac{1 + R'_1}{1 - R'_1}.$$

Then the values making $p(\omega)$ a maximum are:

$$\hat{\xi} = \bar{X}, \qquad \hat{\eta} = \bar{Y}, \qquad \sigma_Y^2 = \tfrac{1}{2}s_X^2(\hat{\gamma} + u)$$
$$\hat{\xi}' = \bar{X}', \qquad \hat{\eta}' = \bar{Y}', \qquad \sigma_Y'^2 = \tfrac{1}{2}s_X'^2(\hat{\gamma} + u')$$

and $\hat{\gamma}$ is the positive root of the equation

$$(N + N')\gamma^2 - (N - N')(u - u')\gamma - (N + N')uu' = 0$$

or

$$\hat{\gamma} = \frac{(N - N')(u - u') + \sqrt{(N - N')^2(u - u')^2 + 4(N + N')uu'}}{2(N + N')} \tag{11}$$

$$= \gamma_1, \text{ say.}$$

Then

$$p(\omega \text{ max}) = \left(\frac{1}{2\pi}\right)^{N+N'} \left[\frac{2\sqrt{\gamma_1}}{(\gamma_1+u)s_X^2}\right]^N \left[\frac{2\sqrt{\gamma_1}}{(\gamma_1+u')s_X'^2}\right]^{N'} e^{-N-N'},$$

and the likelihood ratio criterion for the hypothesis H_1' is

$$\lambda = \frac{p(\omega \text{ max})}{p(\Omega \text{ max})} = \left[\frac{2\sqrt{\gamma_1}\, s_Y}{(\gamma_1+u)s_X}\right]^N \left[\frac{2\sqrt{\gamma_1}\, s_Y'}{(\gamma_1+u')s_X}\right]^{N'} = \left[\frac{2\sqrt{\gamma_1 u}}{\gamma_1+u}\right]^N \left[\frac{2\sqrt{\gamma_1 u'}}{\gamma_1+u'}\right]^{N'}. \tag{12}$$

For H_2', the values the parameters to make $p(\omega)$ a maximum are:

$$\hat{\eta} = \bar{Y}, \qquad \hat{\sigma}_X^2 = \frac{1}{N}\Sigma X^2 \qquad \hat{\sigma}_Y = s_Y$$

$$\hat{\eta}' = \bar{Y}', \qquad \hat{\sigma}_X'^2 = \frac{1}{N'}\Sigma X'^2 \qquad \hat{\sigma}_Y' = s_Y'.$$

Thus

$$p(\Omega \text{ max}) = \left(\frac{1}{2\pi}\right)^{N+N'} \frac{\sqrt{NN'}}{(\Sigma X^2)^{N/2}(\Sigma X'^2)^{N'/2} s_Y^N s_Y'^{N'}} e^{-N-N'}.$$

Similarly, if we write

$$R_2 = \frac{2\gamma s_1 s_2 - \frac{1}{2}(\bar{x}_1-\bar{x}_2)^2}{s_1^2+s_2^2+\frac{1}{2}(\bar{x}_1-\bar{x}_2)^2}, \qquad R_2' = \frac{2\gamma' s_1' s_2' - \frac{1}{2}(\bar{x}_1'-\bar{x}_2')^2}{s_1'^2+s_2'^2+\frac{1}{2}(\bar{x}_1'-\bar{x}_2')^2},$$

and

$$v = \frac{Ns_Y^2}{\Sigma X^2} = \frac{s_Y^2}{s_X^2+\bar{x}^2} = \frac{1+R_2}{1-R_2}, \qquad v' = \frac{Ns_Y'^2}{\Sigma X'^2} = \frac{1+R_2'}{1-R_2'},$$

the values to make $p(\omega)$ a maximum are:

$$\hat{\eta} = \bar{Y}, \qquad \sigma_Y^2 = \frac{1}{2N}\Sigma X^2(\hat{\gamma}+v)$$

$$\hat{\eta}' = \bar{Y}', \qquad \sigma_Y'^2 = \frac{1}{2N'}\Sigma X'^2(\hat{\gamma}+v)$$

$$\hat{\gamma} = \frac{(N-N')(v-v') + \sqrt{(N-N')^2(v-v')^2 + 4(N+N')^2uu'}}{2(N+N')} = \gamma_2, \text{ say.} \tag{13}$$

Then

$$p(\omega \text{ max}) = \left(\frac{1}{2\pi}\right)^{N+N'} \left[\frac{2N\sqrt{\gamma_2}}{(\gamma_2+v)\Sigma X^2}\right]^N \left[\frac{2N'\sqrt{\gamma_2}}{(\gamma_2+v')\Sigma X'^2}\right]^{N'},$$

and the likelihood ratio criterion for the hypothesis H_2' is

$$\lambda_2 = \frac{p(\omega \max)}{p(\Omega \max)} = \left[\frac{2\sqrt{N\gamma_2}\, s_Y}{(\gamma_2 + v)\sqrt{\Sigma X^2}}\right]^N \left[\frac{2\sqrt{N'\gamma_2}\, s_Y'}{(\gamma_2 + v')\sqrt{\Sigma X'^2}}\right]^{N'}$$

(14)

$$= \left[\frac{2\sqrt{\gamma_2 v}}{\gamma_2 + v}\right]^N \left[\frac{2\sqrt{\gamma_2 v'}}{\gamma_2 + v'}\right]^{N'}.$$

The case $N = N'$. The above criteria λ_1 and λ_2 cannot in general be expressed simply, but when $N = N'$, by (11) and (13)

$$\gamma_1 = \sqrt{uu'}, \qquad \gamma_2 = \sqrt{vv'},$$

and

$$\lambda_1 = \left[\frac{4\sqrt{uu'}}{(\sqrt{u} + \sqrt{u'})^2}\right]^N, \qquad \lambda_2 = \left[\frac{4\sqrt{vv'}}{(\sqrt{v} + \sqrt{v'})^2}\right]^N,$$

or we may express as monotonic functions of λ_1 and λ_2,

$$L_1 = \lambda_1^{2/(N+N')} = \lambda_1^{1/N} = \frac{4}{\left(\sqrt[4]{\frac{u}{u'}} + \sqrt[4]{\frac{u'}{u}}\right)^2}, \tag{15}$$

$$L_2 = \lambda_2^{1/N} = \frac{4}{\left(\sqrt[4]{\frac{v}{v'}} + \sqrt[4]{\frac{v'}{v}}\right)^2}. \tag{16}$$

Thus, λ's, L's, or their functions $\frac{u}{u'}$, $\frac{v}{v'}$, may be used as the criteria in the present case.

Furthermore, if we introduce,

$$z = \tfrac{1}{2}\log u, \quad \text{and} \quad z' = \tfrac{1}{2}\log u', \tag{17}$$

we have

$$\tfrac{1}{2}(z - z') = \tfrac{1}{4}\log\frac{u}{u'} \quad \text{or} \quad \sqrt[4]{\frac{u}{u'}} = e^{\frac{1}{2}(z-z')}.$$

Thus L_1 can be written in terms of z and z'

$$L_1 = 4/(e^{\frac{1}{2}(z-z')} + e^{-\frac{1}{2}(z-z')}) = 1/\cosh^2 \tfrac{1}{2}(z - z') = \operatorname{sech}^2 \tfrac{1}{2}(z - z'), \tag{18}$$

and $z - z' = w$, say, may be used also as a criterion for H_1.

We shall now proceed to obtain the distributions of some of these statistics.

4. **The distributions of u/u' and v/v'.** Since Ns_Y^2/σ_Y^2 and Ns_X^2/σ_X^2 have independently the χ^2 distribution with $N - 1$ degrees of freedom,

$$u = \frac{s_Y^2}{s_X^2} = \frac{\sigma_Y^2 \chi_2^2}{\sigma_X^2 \chi_1^2} = \frac{\gamma \chi_2^2}{\chi_1^2}$$

and u/γ has the F distribution with degrees of freedom $f_1 = N - 1, f_2 = N - 1$.

Similarly, $u'/\gamma' = \chi_2'^2/\chi_1'^2$ has the F distribution with the same numbers of degrees of freedom (since $N = N'$, in the present case).

If the hypothesis H_1' is true (i e., $\gamma = \gamma'$)

$$\frac{u}{u'} = \frac{\chi_2^2\chi_1'^2}{\chi_1^2\chi_2'^2} = \frac{\theta_1'\theta_2}{\theta_1\theta_2'} = \frac{z_1}{z_2}, \tag{19}$$

where $\theta_i(-\frac{1}{2}\chi_i^2)$ or θ_1' is distributed as

$$\frac{1}{\Gamma(a_i)}\theta_i^{a_i-1}e^{-\theta_i}\,d\theta_i, \tag{20}$$

with $a_i = \frac{1}{2}(N-1)$, and $z_1(=\theta_1'\theta_2)$, $z_2(=\theta_1\theta_2')$ follow independently the Wilks' z-distribution, [14], which we shall study in detail for the present case

Distribution of z when p = 2. Consider

$$z = B\theta_1\theta_2\cdots\theta_p\,.$$

Wilks has succeeded in integrating the distribution of z for the case $p = 2$ for special values of a's, e.g , $a_1 = \frac{1}{2}(N-1)$, $a_2 = \frac{1}{2}(N-2)$. Now we want the distribution of z when $p = 2$ and for any values of a, and then for $a_1 = a_2 = \frac{1}{2}(N-1)$.

By (20) the joint distribution of θ_1 and θ_2 is

$$\frac{1}{\Gamma(a_1)\Gamma(a_2)}\theta_1^{a_1-1}e^{-\theta_1}\theta_2^{a_2-1}e^{-\theta_2}\,d\theta_1\,d\theta_2\,.$$

Applying the transformation $z = B\theta_1\theta_2$, $v_1 = \theta_1$, the joint distribution of v_1 , z is

$$\frac{1}{\Gamma(a_1)\Gamma(a_2)}v_1^{a_1-1}e^{-v_1}\left(\frac{z}{Bv_1}\right)^{a_2-1}e^{-z/Bv_1}\,\frac{dv_1\,dz}{Bv_1}\,.$$

Integrating v_1 from $v_1 = 0$ to $v_1 = \infty$, we have the distribution of z, viz.,

$$\frac{z^{a_2-1}\,dz}{B^{a_2}\Gamma(a_1)\Gamma(a_2)}\int_0^\infty v^{a_1-a_2-1}e^{-v_1-z/Bv_1}\,dv_1. \tag{21}$$

In order to evaluate the integral of (20), consider the transformation $v_1 = y^2$, $dv_1 = 2y\,dy$, we have

$$I_0 = 2\int_0^\infty y^{2(a_1-a_2)-1}e^{-y^2-z/By^2}\,dy \tag{22}$$

To evaluate I_0 for any a's, by putting $y = 1/x$, $dy = -dx/x^2$, we have

$$I_0 = 2\int_0^\infty \frac{e^{-zx^2/B-1/x^2}}{x^{2(a_1-a_2)+1}}\,dx. \tag{23}$$

Consider

$$\frac{\Gamma(a_1-a_2+\frac{1}{2})}{x^{2(a_1-a_2)+1}} = \int_0^\infty e^{-x^2y}y^{a_1-a_2-\frac{1}{2}}\,dy. \tag{24}$$

Then

$$I_0\Gamma(a_1 - a_2 + \tfrac{1}{2}) = 2\int_0^\infty e^{-(zx^2/B+1/x^2)}\,dx \int_0^\infty e^{-x^2y} y^{a_1-a_2-\frac{1}{2}}\,dy$$

$$= 2\int_0^\infty y^{a_1-a_2-\frac{1}{2}}\,dy \int_0^\infty e^{-[(z/B+y)x^2+1/x^2]}\,dx$$

$$= \sqrt{\pi}\int_0^\infty e^{-2\sqrt{z/B+y}}\, y^{a_1-a_2-\frac{1}{2}}\,\frac{dy}{\sqrt{z/B+y}}.$$

Since by the substitution $\sqrt{\frac{z}{B}+y} = \sqrt{\frac{z}{B}} + y$ or $y = x^2 + 2\sqrt{\frac{z}{B}}\,x$, $dy = 2\left(x + \sqrt{\frac{z}{B}}\right)dx$ and therefore

$$I_0\Gamma(a_1 - a_2 + \tfrac{1}{2}) = 2\sqrt{\pi}\int_0^\infty e^{-2(\sqrt{z/B}+x)}\left(x^2 + 2x\sqrt{\frac{z}{B}}\right)^{a_1-a_2-\frac{1}{2}}dx,$$

$$I_0 = \frac{\sqrt{\pi}\,e^{-2\sqrt{z/B}}}{\Gamma(a_1 - a_2 + \frac{1}{2})}\int_0^\infty e^{-2(\sqrt{z/B}+x)}\left(x^2 + 2x\sqrt{\frac{z}{B}}\right)^{a_1-a_2-\frac{1}{2}}dx. \tag{25}$$

Hence, z is distributed as

$$\frac{2\sqrt{\pi}\,z^{a_2-1}e^{-2\sqrt{z/B}}}{B^{a_2}\Gamma(a_1)\Gamma(a_2)\Gamma(a_1 - a_2 + \frac{1}{2})}\int_0^\infty e^{-2x}\left(2\sqrt{\frac{z}{B}} + x\right)^{a_1-a_2-\frac{1}{2}} x^{a_1-a_2-\frac{1}{2}}\,dx. \tag{26}$$

We infer from this distribution that when $2(a_1 - a_2)$, i.e., the difference of degrees of freedom, is odd, the integral can be expressed as a terminated series; but for even values of $2(a_1 - a_2)$, the series is infinite.

When $B = \frac{1}{A}$, $a_1 = \frac{1}{2}(N-1)$, $a_2 = \frac{1}{2}(N-2)$, (26) is reduced to

$$\frac{\sqrt{\pi}\,A^{a_2}z^{a_2-1}e^{-2\sqrt{Az}}}{\Gamma(a_1)\Gamma(a_2)}, \tag{27}$$

which is Wilks' ξ distribution, [15], for $p = 2$.

When $B = 1$ and $a_1 = a_2 = \frac{1}{2}(N-1)$, it becomes

$$\frac{2\sqrt{\pi}\,z^{a_2-1}e^{-2\sqrt{z}}}{\Gamma(a_1)\Gamma(a_2)}\int_0^\infty e^{-2x}(2\sqrt{z} + x)^{-\frac{1}{2}}x^{-\frac{1}{2}}\,dx, \tag{28}$$

which is the distribution of z involved in (19).

Since (28) can apparently not be simplified, I have been unable thus far to find in manageable form the distribution of the ratio z_1/z_2 and therefore of u/u' in this case. However, it would be simpler to use the alternative criterion $w = z - z'$ for the hypothesis H_1. The distribution of w will be taken up in a later section.

The distribution of v/v': Since Ns_Y^2/σ_Y^2 and $\Sigma X^2/\sigma_X^2$ have independently the χ^2 distribution with $N-1$ and N degrees of freedom respectively, therefore,

$$v = \frac{NS_Y^2}{\Sigma X^2} = \frac{\sigma_Y^2 \chi_2^2}{\sigma_X^2 \chi_1^2} = \frac{\gamma \chi_2^2}{\chi_1^2},$$

and $\frac{v}{\gamma} \Big/ \frac{N-1}{N}$ has the F-distribution with $f_1 = N-1$ degrees of freedom and $f_2 = N$.

Similarly $\frac{v'}{\gamma'} \Big/ \frac{N-1}{N}$ has the F-distribution with degrees of freedom f_1 and f_2 as above.

If the hypothesis H_2 is true (i.e., $\gamma = \gamma'$),

$$\frac{v}{v'} = \frac{\chi_2^2 \chi_1'^2}{\chi_1^2 \chi_2'^2} = \frac{\theta_1' \theta_2}{\theta_1 \theta_2'} = \frac{z_1}{z_2},$$

where each θ_i is distributed as in (19), but with $a_1 = \frac{1}{2}N$ and $a_2 = \frac{1}{2}(N-1)$. We can infer from (27) that $t_1 = 4\sqrt{z_1}$ and $t_2 = 4\sqrt{z_2}$ have independently the χ^2-distribution each with $4a_2$ or $2(N-1)$ degrees of freedom, and $t_1/t_2 = \sqrt{z_1/z_2} = \sqrt{v/v'}$ follows the F-distribution with degrees of freedom $f_1 = f_2 = 2(N-1)$. The 5% and 1% points of the $F = v/v'$ may be obtained from Snedecor's table ([12], p. 174).

5. **The distribution of $y = \log z$.** Wald [13] has suggested that the distribution of $z = B\theta_1\theta_2 \cdots \theta_p$ for any a_i's ($i = 1, \cdots, p$) may also be obtained indirectly with the aid of the characteristic function. A similar method has been applied in a recent paper by Wald and Brookner [14]. Consider the transformation

$$y = \log t = \log B\theta_1\theta_2 \cdots \theta_p. \tag{29}$$

The characteristic function of y is

$$\begin{aligned} \varphi_y(t) &= E(e^{ty}) = E\{(B\theta_1\theta_2 \cdots \theta_p)^t\} \\ &= \frac{B^t \Gamma(a_1+t)\Gamma(a_2+t) \cdots \Gamma(a_p+t)}{\Gamma(a_1)\Gamma(a_2) \cdots \Gamma(a_p)}. \end{aligned} \tag{30}$$

Thus the distribution $f(y)\,dy$ is given by

$$f(y) = \frac{1}{2\pi i} \int_{-\infty}^{i\infty} e^{-ty} \varphi_y(t)\,dt = \frac{1}{2\pi i} \int_{-\infty}^{i\infty} B^t e^{-ty} \prod_{i=1}^{p} \frac{\Gamma(a_i+t)}{\Gamma(a_i)}\,dt. \tag{31}$$

Without loss of generality, we may take $a_1 \geqq a_2 \geqq \cdots \geqq a_p > 0$ and let $a_p + t = -t'$, then

$$f(y) = \frac{c_p}{2\pi i} \int_{a_p - i\infty}^{-a_p + i\infty} e^{yt'} B^{-t'} \prod_{i=1}^{p} \Gamma(a_i - a_p - t')\,dt', \tag{32}$$

where $c_p = e^{a_p y} B^{-a_p} \Big/ \prod_{i=1}^{p} \Gamma(a_i)$.

The integration can be carried out by the method of residue along the contour C, bounded by the line $x = -a_p$ and that part of the circle with center at origin and radius r, which lies to the right of the line $x = -a_p$. The integral of the function $e^{t'v}B^{-t'}\prod_{i=1}^{p}\Gamma(a_i - a_p - t')$ along the arc converges to zero as the radius of the circle tends to infinity (Kullback, [8]). Hence the integrals along the vertical line $x + a_p = 0$ and along the closed contour C are equal. Then we may write

$$f(y) = -\frac{c_p}{2\pi i}\int_C e^{vt'}B^{-t'}\prod_{i=1}^{p}\Gamma(a_i - a_p - t')\,dt', \tag{33}$$

and its value is c_p times the sum of the residues at the poles within the contour C.

For the present purpose, $p = 2$, we have

$$f(y) = \frac{c_2}{2\pi i}\int_{-a_2+i\infty}^{-a_2+i\infty} e^{vt'}\Gamma(a_1 - a_2 - t')\Gamma(-t')\,dt'. \tag{34}$$

We shall study the integral of (34) in more detail in the following cases:

(i) $a_1 - a_2 = \frac{1}{2}$. By the duplication formula

$$\Gamma(\tfrac{1}{2} - t')\Gamma(-t') = 2^{1+2t'}\sqrt{\pi}\,\Gamma(-2t'),$$

and the function

$$\Gamma(-2t') = \lim_{N\to\infty}\frac{N!N^{-2t'}}{(-2t')(-2t'+1)\cdots(-2t'+N)},$$

has simple poles at the points $0, \frac{1}{2}, 1, 3/2, \ldots$. The residue at $t' = m/2$, where m is zero or a positive integer, is $(-1)^{m+1}/2\cdot m!$ and (34) becomes

$$\begin{aligned} f(y) &= \sqrt{\pi}\,c_2\left(1 - 2e^{\frac{1}{2}v} + \frac{1}{2!}2^2e^{v} - \frac{1}{3!}2^3e^{3v/2} + \cdots\right) \\ &= \sqrt{\pi}\,c_2e^{-2e^{\frac{1}{2}v}}. \end{aligned} \tag{35}$$

The distribution of $z = e^{v}$ is

$$\frac{2\sqrt{\pi}\,z^{a_2-1}e^{-2\sqrt{z}}}{\Gamma(a_1)\Gamma(a_2)}\,dz. \tag{27 bis}$$

(ii) $a_1 - a_2 = m + \frac{1}{2}$. The function

$$\begin{aligned} \Gamma(a_1 - a_2 - t')\Gamma(-t') &= (m - \tfrac{1}{2} - t')(m - \tfrac{3}{2} - t')\cdots(\tfrac{1}{2} - t')\Gamma(\tfrac{1}{2} - t')\Gamma(-t') \\ &= 2^{1+2t'}\sqrt{\pi}(m - \tfrac{1}{2} - t')(m - \tfrac{3}{2} - t')\cdots(\tfrac{1}{2} - t')\Gamma(-2t') \end{aligned}$$

has simple poles at 0, m, $m + \frac{1}{2}$, $m + 1$, $\cdots$, and

$$f(y) = \sqrt{\pi}\, c_2 \left[\frac{(2m-1)!}{2^{2m-1}(m-1)!} - \frac{1}{2^m(2m)} (2^2 e^y)^m + \frac{1}{2^m(2m+1)} (2^2 e^y)^{m+\frac{1}{2}} \right.$$

$$\left. - \frac{1}{2^m(2m+2)\cdot 2\cdot 1} (2^2 e^y)^{m+1} + \cdots \right]$$

$$= \sqrt{\pi}\, c_2 \left[\frac{(2m-1)!}{2^{2m-1}(m-1)!} - \frac{1}{2^m} \sum_{\gamma=0}^{\infty} \frac{1}{(2m+\gamma)\gamma!} (2^2 e^y)^{m+\gamma/2} \right].$$

This agrees with the expansion of (26) when we put $a_1 - a_2 - \frac{1}{2} = m$.

(iii) $a_1 - a_2 = 0$. The function

$$[\Gamma(-t')]^2 = \lim_{N\to\infty} \frac{(N!)^2 N^{-2t'}}{(-t')^2(-t'+1)^2 \cdots (-t'+N)^2},$$

has poles of the second order at the points 0, 1, 2, 3, $\cdots$ and

$$f(y) = c_2 \sum_{\gamma=0}^{\infty} \frac{d}{dt'} \{(t'-\gamma)^2 e^{t'y} [\Gamma(-t')]^2\}_{t'=\gamma}$$

(iv) $a_1 - a_2 = m$. The function

$$\Gamma(m - t')\Gamma(-t') = (m-1-t')(m-2-t') \cdots (1-t')(-t')[\Gamma(-t')]^2,$$

has finite simple poles at 1, 2, $\cdots$, $m-1$ and poles of the second order at m, $m+1$, $\cdots$, and

$$f(y) = c_2 \sum_{\gamma=0}^{m-1} \{(t-\gamma)e^{t'y}\Gamma(m-t')\Gamma(-t')\}_{t'=\gamma}$$

$$+ c_2 \sum_{\gamma=m}^{\infty} \left\{ \frac{d}{dt'} (t'-\gamma)^2 e^{t'y} \Gamma(m-t')\Gamma(-t') \right\}_{t'=\gamma}.$$

6. **The distribution of $w = z - z'$ or $\psi = \cosh w$.** Since the distribution of u is given in [7] as

$$\frac{1}{\gamma \mathrm{B}[\frac{1}{2}(N-1), \frac{1}{2}(N-1)]} \left(\frac{u}{\gamma}\right)^{\frac{1}{2}N-3} \left(1 + \frac{u}{\gamma}\right)^{-(N-1)} du, \tag{39}$$

therefore, by transformation (17), we have that the distribution of z for a given $\zeta = \frac{1}{2}\log\gamma = \frac{1}{2}\log\frac{1+\rho}{1-\rho}$ is

$$\frac{1}{\mathrm{B}\left(\frac{1}{2}, \frac{n}{2}\right)} \operatorname{sech}^n (z-\zeta)\, dz, \tag{40}$$

where $n = N - 1$. The distribution of z has been given by R. A. Fisher [3] for $n = 1$ and by Delury [2]. Similarly, the distribution of z' for a given ζ' is

$$\frac{1}{\mathrm{B}\left(\frac{1}{2}, \frac{n'}{2}\right)} \operatorname{sech}^{n'}(z' - \zeta')\, dz', \tag{41}$$

where $n' = N' - 1$.

In case $n = n'$, the joint distribution of z and z' for a given common ζ is

$$C \operatorname{sech}^{n}(z - \zeta) \operatorname{sech}^{n}(z' - \zeta)\, dz\, dz' = \frac{C\, dz\, dz'}{\cosh^{n}(z - \zeta) \cosh^{n}(z' - \zeta)}, \tag{42}$$

where $1/C = \left[\mathrm{B}\left(\frac{1}{2}, \frac{n}{2}\right)\right]^2$.

By the transformation $\bar{z} = \frac{1}{2}(z + z')$, $w = z - z'$, we have the joint distribution of $\bar{z}$ and w,

$$\frac{C\, d\bar{z}\, dw}{[\cosh^{n}(z - \zeta) \cosh^{n}(z' - \zeta)]} = \frac{2^{n} C\, d\bar{z}\, dw}{[\cosh 2(\bar{z} - \zeta) + \cosh w]^{n}}. \tag{43}$$

Integrating with respect to $\bar{z}$ from $-\infty$ to ∞, we have

$$\begin{aligned} &2^{n} C\, dw \int_{-\infty}^{\infty} \frac{d\bar{z}}{[\cosh 2(\bar{z} - \zeta) + \cosh w]^{n}} \\ &\qquad = 2^{n} C\, dw \int_{0}^{\infty} \frac{2\, d\bar{z}}{[\cosh 2(\bar{z} - \zeta) + \cosh w]^{n}} \\ &\qquad = 2^{n} C\, dw I_{n}, \text{ say.} \end{aligned} \tag{44}$$

Applying the transformation $\phi = 2(\bar{z} - \zeta)$, $\psi = \cosh w$, the integral of (34) becomes

$$I_{n} = \int_{0}^{\infty} \frac{d\phi}{(\cosh \phi + \psi)^{n}}.$$

Substituting $\cosh \phi + \psi = \dfrac{1 + \psi}{\theta}$, we have

$$\begin{aligned} I_{n} &= \int_{0}^{1} \left(\frac{\theta}{1 + \psi}\right)^{n} \frac{1}{\theta} \frac{d\theta}{\sqrt{\left(1 - \dfrac{\psi - 1}{\psi + 1}\theta\right)(1 - \theta)}} \\ &= \frac{1}{(\psi + 1)^{n}} \int_{0}^{1} \theta^{n-1} (1 - \theta)^{-\frac{1}{2}} \left(1 - \frac{\psi - 1}{\psi + 1}\theta\right)^{-\frac{1}{2}} d\theta. \end{aligned} \tag{45}$$

Comparing (35) with the hypergeometric function

$$I = \int_{0}^{1} \theta^{b-1} (1 - \theta)^{c-b-1} (1 - \theta x)^{-a}\, d\theta = \frac{\Gamma(b)\Gamma(c - b)}{\Gamma(c)} F(a, b, c, x), \tag{46}$$

we have $b = n$, $c - b = \frac{1}{2}$, $a = \frac{1}{2}$, and therefore (35) can be expressed in terms of a hypergeometric series as

$$I_n = \frac{\Gamma(n)\Gamma(\frac{1}{2})}{\Gamma(n+\frac{1}{2})} \frac{1}{(\psi+1)^n} F\left(\tfrac{1}{2}, n, n+\tfrac{1}{2}, \frac{\psi-1}{\psi+1}\right). \tag{47}$$

The series (37) is convergent since $\frac{\psi-1}{\psi+1}$ is less than unity. Thus the distribution of w, from (34), is

$$\frac{2^n C\Gamma(n)\Gamma(\frac{1}{2})}{\Gamma(n+\frac{1}{2})} \frac{1}{(\cosh w+1)^n} F\left(\tfrac{1}{2}, n, n+\tfrac{1}{2}, \frac{\cosh w-1}{\cosh w+1}\right) dw, \tag{48}$$

and the distribution of $\psi = \cosh w$ is

$$\frac{2^{n+1} C\Gamma(n)\Gamma(\frac{1}{2})}{\Gamma(n+\frac{1}{2})} \frac{1}{(\psi+1)^{n+\frac{1}{2}}(\psi-1)^{\frac{1}{2}}} F\left(\tfrac{1}{2}, n, n+\tfrac{1}{2}, \frac{\psi-1}{\psi+1}\right) d\psi. \tag{49}$$

We notice that the distribution of ψ expressed in (39) is very similar to the r-distribution expressed in terms of hypergeometric series, except that in the first case the argument is $\frac{\psi-1}{\psi+1}$, while in the second case it is $\frac{1-p}{1+p}$ where $p = \rho r$. Hotelling [5] has obtained a very rapidly convergent hypergeometric series for the distribution of the correlation coefficient since $|p| < 1$. But for the distribution of ψ, we cannot obtain a more rapidly convergent series than (39), since the values of ψ lie between 1 and ∞.

7. **Summary and remark.** Two hypotheses concerning the comparison of correlation coefficients of two samples from bivariate normal populations have been considered. The appropriate test criteria for each hypothesis have been derived by the use of a transformation of the variates. The distributions of certain of the criteria have been obtained in the special case where $N = N'$. Incidentally the distribution of Wilks' z for $p = 2$ and any values of a_1 and a_2 has been derived.

Again though we assume throughout the paper that $\sigma_1 = \sigma_2$ and $\sigma_1' = \sigma_2'$, the tests can be generalized to fit the case where the ratios $\sigma_1/\sigma_2 = k$, $\sigma_1'/\sigma_2' = k'$ are known, but are different from unity. In the latter case we can apply the transformation

$$y_1 = w_1 x_1, \qquad y_2 = w_2 x_2;$$
$$y_1' = w_1' x_1', \qquad y_2' = w_2' x_2';$$

where

$$w_1 k_1 = w_2 k_2 = 1, \qquad w_1' k_1' = w_2' k_2' = 1,$$

so that after transformation the variances of each pair of y's are equal.

The writer is deeply indebted to Professor Harold Hotelling and Dr. Abraham Wald for their advice and suggestions in the preparation of this paper.

REFERENCES

[1] F A. Brander, *Biometrika*, Vol 35 (1933), p 102.
[2] D. B. Delury, *Annals of Math. Stat.*, Vol 9 (1938), p. 145.
[3] R. A. Fisher, *Metron*, Vol. 1 (1921), p. 3.
[4] H. Hotelling, *Biometrika*, Vol. 26 (1936), p. 321.
[5] H. Hotelling, Lectures delivered at Columbia University (1940–41).
[6] H. Hotelling, *Annals of Math. Stat.*, Vol. 11 (1940), p. 271.
[7] C. T. Hsu, *Annals of Math. Stat.*, Vol. 11 (1940), p. 410.
[8] S. Kullback, *Annals of Math. Stat.*, Vol. 5 (1934), p. 263.
[9] J Neyman and E S. Pearson, *Biometrika*, Vol. 20 (1928), p. 175.
[10] J. Neyman and E. S. Pearson, *Bull. Acad. Polonaise Sci. Lettres A* (1931), p. 460.
[11] E. S. Pearson and J. Neyman, *Bull. Acad. Polonaise Sci. Lettres A* (1930), p 73.
[12] G. W. Snedecor, *Statistical Methods*, Collegiate Press, Ames, Iowa, 1937.
[13] A. Wald, Lectures delivered at Columbia University (1939–40).
[14] A. Wald and R. J. Brookner, *Annals of Math. Stat* , Vol. 12 (1941), p. 137.
[15] S. S. Wilks, *Biometrika*, Vol. 24 (1932), p. 471.

ON RANDOMNESS IN ORDERED SEQUENCES

By L. C. Young

Westinghouse Electric and Manufacturing Company

It is frequently desirable to examine an ordered sequence of measurements for the presence of non-random variability, concern over any particular type of variability being limited Unless the sequence is one containing replicated observations, current methods of analysis often restrict an investigation to tests for specific forms of variability, such as particular orders of regression and periodicity. In order to simulate replication, arbitrary grouping of data is occasionally used and followed by some test of variance; this practice, however, is likely to add an element of bias to the investigation

Under these conditions, it would be convenient to have the means of testing a series for the presence of general regression, before proceeding to test for that of a specific type. It is the purpose of this paper to present, as briefly as possible, a statistic designed for this preliminary type of examination, and to demonstrate its application.

If a given sequence of measurements be denoted by

$$X_1, X_2, \cdots, X_n$$

then the magnitude of

$$C = 1 - \frac{\sum_1^{n-1} (X_i - X_{i+1})^2}{2 \sum_1^n (X_i - \bar{X})^2},$$

will be dependent upon the arrangement of the n observations upon which it is based. C will have $n!$ possible values for a given sample, corresponding to the number of permutations of n items.

1. Moments of the distribution of C in terms of the moments of a finite sequence. Writing C in terms $x_1, \cdots, x_n$, representing the deviations of $X_1, \cdots, X_n$ from their sample mean of n measurements,

$$C = 1 - \frac{\sum_1^{n-1} (x_i - x_{i+1})^2}{2 \sum_1^n x_i^2}$$

$$= \frac{x_1^2 + x_n^2 + 2 \sum_1^{n-1} x_i x_{i+1}}{2 \sum_1^n x_i^2},$$

In order to find the mean value of C for a given sample, it must be summed over all values obtained from the $n!$ permutations of the measurements.

Dealing with the numerator alone of the expression given above:

$$\sum_p \left[x_1^2 + x_n^2 + 2 \sum_1^{n-1} x_i x_{i+1} \right] = \sum_p x_1^2 + \sum_p x_n^2 + 2 \sum_p \sum_1^{n-1} x_i x_{i+1},$$

where Σ_p denotes summation over the $n!$ permutations.

There are n values of x_i, and $n!$ arrangements. Each value x_i is x_1 in $(n-1)!$ of the arrangements: the same reasoning applies to x_n. The first two terms of the summation, therefore, will be

$$\sum_p x_1^2 = \sum_p x_n^2 = (n-1)! \sum_1^n x_i^2.$$

With regard to the third term, there are $2(n-1)$ of such cross-products for each arrangement. Since the summation is taken over $n!$ arrangements, $x_i x_k$ will be different than $x_k x_i$, and should be considered a separate term. Each crossproduct term, therefore, must occur $\frac{2(n!)(n-1)}{n(n-1)}$ times throughout the $n!$ arrangements, since there are $n(n-1)$ possible cross-products among n different items The third term, then, will be

$$2 \sum_p \left(\sum_1^{n-1} x_i x_{i+1} \right) = 2(n-1)! \sum_1^n \sum_1^{n-1} x_i x_k = -2(n-1)! \sum_1^n x_i^2,$$

from which it may be seen that the mean value of C is zero for any sample.

The same method may be applied in order to find the second and higher moments of C. Squaring the numerator of the expression and expanding,

$$\sum_p \left[x_1^2 + x_n^2 + 2 \sum_1^{n-1} x_i x_{i+1} \right]^2$$
$$= \sum_p \left[x_1^4 + x_n^4 + 2x_1^2 x_n^2 + 4x_1^2 \sum_1^{n-1} x_i x_{i+1} + 4x_n^2 \sum_1^{n-1} x_i x_{i+1} + 4 \left(\sum_1^{n-1} x_i x_{i+1} \right)^2 \right].$$

Performing the summation Σ_p term by term we obtain

$$\frac{\sum_p \left[x_1^2 + x_n^2 + 2 \sum_1^{n-1} x_i x_{i+1} \right]^2}{n!} = \frac{2(2n-3) \left(\sum_1^n x_i^2 \right)^2 - 2n \sum_1^n x_i^4}{n(n-1)},$$

whence the second moment of C for any sample is given by

$$M_2 = \frac{2n - 3 - m_4/m_2^2}{2n(n-1)},$$

where m_2 and m_4 are the second and fourth moments, respectively, of the n observations about their mean.

In like manner, the third and fourth moments of the distribution of C for a given sample of n observations are found to be

$$M_3 = \frac{-6 + 4(n-3)\frac{m_3^2}{m_2^3} + 9\frac{m_4}{m_2^2} - 3\frac{m_6}{m_2^3}}{4n(n-1)(n-2)},$$

$$M_4 = \frac{1}{8n^3(n-1)(n-2)(n-3)}\left[24n^2(n-3)^2 - 48n(4n-9)\frac{m_3^2}{m_2^3}\right.$$
$$- 24n(3n^2 - 17n + 27)\frac{m_4}{m_2^2} + (8n^3 - 45n^2 - 23n + 210)\frac{m_4^2}{m_2^4}$$
$$+ 16(2n^2 + 5n - 21)\frac{m_5 m_3}{m_2^4} + 4(17n^2 - 37n + 42)\frac{m_6}{m_2^3}$$
$$\left. - (7n^2 + 13n - 6)\frac{m_8}{m_2^4}\right].$$

2. Distribution of C for samples drawn from a normal universe. The first four moments of the distribution of C for samples drawn from a given population may be derived from the above formulae by substituting the mean values of $\frac{m_3^2}{m_2^3}, \frac{m_4}{m_2^2}$, etc. of samples from such a population. For normal samples containing n observations, for example, the following mean values apply, as obtained by the method presented by R. A. Fisher [1, 2]:

$$\frac{m_3^2}{m_2^3} = \frac{6(n-2)}{(n+1)(n+3)},$$

$$\frac{m_4}{m_2^2} = \frac{3(n-1)}{(n+1)},$$

$$\frac{m_4^2}{m_2^4} = \frac{3(3n^3 + 23n^2 - 63n + 45)}{(n+1)(n+3)(n+5)},$$

$$\frac{m_5 m_3}{m_2^4} = \frac{60(n-1)(n-2)}{(n+1)(n+3)(n+5)},$$

$$\frac{m_6}{m_2^3} = \frac{15(n-1)^2}{(n+1)(n+3)},$$

$$\frac{m_8}{m_2^4} = \frac{105(n-1)^3}{(n+1)(n+3)(n+5)}.$$

Replacement of the sample moment ratios by the mean values of those ratios for normal samples yields the following moments of C:

$$M_1 = 0, \qquad M_2 = \frac{n-2}{(n-1)(n+1)}, \qquad M_3 = 0,$$

$$M_4 = \frac{3(n^2 + 2n - 12)}{(n-1)(n+1)(n+3)(n+5)}.$$

Compatible results for the case of normal samples have been obtained by Williams [3], using another method.

From the above results, the value of

$$\beta_2 = \frac{3(n^2 + 2n - 12)(n - 1)(n + 1)}{(n - 2)^2(n + 3)(n + 5)},$$

is seen to approach normality as the sample size is increased.

Inasmuch as the distribution of C for normal samples is limited in both directions and is symmetrical, it is apparent that the Pearson Type II distribution may be considered representative. Fitting this curve to the moments given above, the equation of the frequency distribution is given by

$$y = y_0\left(1 - \frac{C^2}{a^2}\right)^m,$$

where

$$m = \frac{(n^4 - n^3 - 13n^2 + 37n - 60)}{2(n^3 - 13n + 24)},$$

$$a^2 = \frac{(n^2 + 2n - 12)(n - 2)}{(n^3 - 13n + 24)},$$

$$y_0 = \frac{\Gamma(2m + 2)}{a \cdot 2^{2m+1}[\Gamma(m + 1)]^2}.$$

The values of β_2 for the distribution, for various values of n, are as follows:

Sample size, n	β_2
5	2.300
10	2.570
15	2.684
20	2.750
25	2.793
50	2.833

Due to the effect of even moments higher than the fourth, the approximation afforded by the Type II curve is not reliable for samples containing less than about eight observations. As the sample size decreases below this limit, the extremes of the C distribution deviate increasingly from the extremes ($\pm a$) of the fitted curve: with such a platykurtic distribution, therefore, the effect upon the lower significance levels vitiates the approximation

Although either β_2 or the theoretical limits of the distribution of C could have been employed as a parameter of the fitted curve, it was considered expedient to use the former. In any case, of course, the advantage to be gained would be in connection only with samples containing few observations (less than eight). The evidence afforded by empirical sampling indicates that use of the limits as a parameter might render the approximation less valid.

In order to facilitate use of the approximate distribution for samples of eight or more observations, the values of C associated with two probability levels are tabulated below in Table I. The ratio of each value of C to its standard error is also shown, to demonstrate the approach to normality. The significance levels recorded exclude 10% and 2% of the area under the curve, respectively. In most practical applications, these will be the 5% and 1% levels, respectively, since only positive values of C exceeding the tabulated value will ordinarily be considered significant. The tabulations were prepared from tables of the function $I_x(p, q)$ [5], where $q = .5$ and $p = m + 1$, with the transformation $x = 1 - \frac{C^2}{a^2}$.

TABLE I

Significance levels of the absolute value of C

Sample size, n	$P = .10$	$C_{.10}/\sigma_c$	$P = .02$	$C_{.02}/\sigma_c$
8	.5088	1 6486	.6686	2.1664
9	.4878	1 6492	.6456	2.1826
10	.4689	1.6494	.6242	2.1958
11	.4517	1.6495	.6044	2.2068
12	.4362	1.6495	.5860	2.2161
13	.4221	1 6495	.5691	2.2241
14	.4092	1 6494	.5534	2.2310
15	.3973	1.6493	.5389	2.2369
16	.3864	1.6492	.5254	2.2423
17	.3764	1.6492	.5128	2.2470
18	3670	1.6491	.5011	2.2513
19	.3583	1 6489	.4900	2.2550
20	.3502	1.6488	.4797	2.2585
21	.3426	1.6488	.4700	2.2616
22	3355	1.6486	.4609	2.2647
23	.3288	1.6485	.4521	2.2676
24	.3224	1.6484	.4440	2.2700
25	.3165	1.6484	.4361	2.2717
Normal ($n = \infty$)		1.6447		2.3262

The distribution of C for normal samples containing 20 or more observations is sufficiently normal, for most practical cases and for the more common significance levels, to permit use of a table of areas under the normal curve, in conjunction with the standard error $\sigma_c = \sqrt{\frac{n-2}{(n-1)(n+1)}}$. The 5% significance levels shown in Table I result, at worst, in a one per cent error of probability estimate, if the normal approximation is used in their place: that is, if 1.6447 times the standard error is used instead of the tabulated significance level, the probability will be .0505 at most, for the values of n which are tabulated.

3. General discussion on the application of C. It may be wondered why the statistic C has been used, rather than the more easily computed statistic $C' = \dfrac{\sum_1^{n-1} (X_i - X_{i+1})^2}{\sum_1^n x_i^2}$. As far as a significance test is concerned, it clearly does not matter which is used, since C and C' are linearly related. However, C may be regarded as symmetrically distributed about 0 in samples from a normal population to within at least four moments. Excessive departure of C from 0 may be taken as indicative of the presence of non-randomness in the series, the actual significance test being based, of course, on the probability of obtaining a departure larger than a given observed one, under the assumption of a random series. Positive values of C, in general, correspond to positive correlation while negative values correspond to negative correlation between successive observations.

There are various ways of detecting non-randomness in a series of observations, such as regression methods, analysis of variance, etc. The use of regression methods implies that we must know in general the type of regression function to be tried. C is a very flexible statistic, on the other hand, for testing the null hypothesis that a series is random, no matter what the alternative hypothesis is A thorough study of C as a statistic for testing the hypothesis of randomness in an ordered series should include a study of the power function of C for hypotheses specifying various types of non-randomness. However, we shall simply appeal to intuition in proposing the statistic C, and forego power function considerations in this note. In practice, the advantage of using C increases with the length of a series: lack of randomness in a single sequence of ten or less observations may ordinarily be detected by regression methods, in fitting a low order polynomial. In a longer sequence of measurements, on the other hand, the presence of complicated regression or of periodicity is often sufficiently obscured by variation to elude detection by any other than a flexible method.

The statistic could be used to advantage in the field of applied statistics, in the investigation not only of variate series but of attribute series as well. For the latter purpose, an effort to tabulate the relationship between the level of significance and the percentage of either attribute would facilitate statistical investigation of random arrangement. A direct application could thus be made to binomially distributed attributes by a scalar assignment (0, 1) to the dichotomy, followed by a procedure similar to that presented above. Similarly, the randomness of vectorial observations could be examined from the viewpoint of arrangement. The common method of treating such problems,—the "random walk method,"—has occasionally been found inadequate in dealing with specific forms of non-random order; this is especially true when the allocable cause of variation has a multi-directional effect.

Needless to say, each of the fields of application considered so briefly above would require development before a routine, efficient method of investigating ordered arrangement could be established. Although probability level tables

have been provided in this paper for C as applied to normal samples, it is quite evident that tables for samples from other parent distributions would be needed for some of the applications mentioned above.

4. An illustration of the use of C. Although one example has already been presented elsewhere [4] in which the distribution developed in Section 2 has been employed, a typical application of the statistic to an example in the field of quality control will be given here in order to illustrate the mechanics of solution. The data presented in Table II represent the percentages of defective product turned out daily, over a period of twenty-four days, by a single workman. The total output each day closely approximates five hundred parts. this fact is brought out to explain the calculation of χ^2 for the observed series of percentages, —it has no bearing upon the use of C.

TABLE II

Percentage of product rejected

Day	%, X	X^2	d^2
1	7.4	54.76	
2	8.8	77.44	1.96
3	11.4	129.96	6.76
4	10.3	106.09	1.21
5	11.9	141.61	2.56
6	12.2	148.84	.09
7	10.0	100.00	4.84
8	8.4	70.56	2.56
9	9.4	88 36	1.00
10	10.9	118.81	2.25
11	9.9	98 01	1.00
12	11.8	139.24	3.61
13	10.0	100.00	3 24
14	8.9	79.21	1.21
15	9.7	94 09	.64
16	9.3	86.49	.16
17	12.0	144.00	7.29
18	12.3	151.29	.09
19	10.3	106.09	4.00
20	8.6	73.96	2.89
21	10.4	108.16	3.24
22	11.1	123.21	.49
23	9.4	88.38	2.89
24	8.2	67.24	1.44
Totals	242.6	2495.82	55.42
		$n\bar{X}^2$ 2452.28	
		$\Sigma x^2 =$ 43.54	

$C = .3636$ (significant) $\chi^2 = 21.518$ (23 degrees of freedom) (not significant).

The value of C derived from the data lies between the two significance levels tabulated in Table I; there is reason to believe that the data are ordered, or non-random. Computation of χ^2, however, has been carried out with the hypothesis that all product was made under the same conditions (i.e. with a percentage defective equal to 10.108%, the mean of the group). The value so obtained is associated with a probability of about $P = .50$: the hypothesis is not disproved by this test. In short, the variability of the twenty-four observations could be considered random if it were not for the order of their arrangement.

REFERENCES

[1] R. A. Fisher, "Moments and product moments of sampling distributions," *Lond. Math. Soc. Proc.* (series 2) 30 (1929), pp. 199–238.

[2] R. A. Fisher, "The moments of the distribution for normal samples of measures of departure from normality," *Roy. Soc. Proc*, A 130 (1930), pp. 16–28

[3] J D. Williams, "Moments of the ratio of the mean square successive difference in samples from a normal universe," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 239-241.

[4] L. C. Young, "A critical appraisal of statistical methods in industrial management," presented at the annual meeting, American Society of Mechanical Engineers (1940)

[5] K. Pearson (Editor), *Tables of the Incomplete Beta-Function*, Biometrika Office, London, 1924.

ON CERTAIN LIKELIHOOD-RATIO TESTS ASSOCIATED WITH THE EXPONENTIAL DISTRIBUTION

By Edward Paulson

Washington, D.C.

Various likelihood-ratio tests and their distributions in samples from a population having the elementary probability law $\frac{1}{\sigma} e^{-(x-B)/\sigma}$, $B \leq x \leq \infty$, have been studied by Neyman and Pearson [1] and Sukhatme [2]. In this note the power functions and the question of bias of several likelihood-ratio tests will be investigated. The exponential distribution appears to be appropriate for dealing with problems involving the intervals of time between events which tend to be random, as for example the interval between consecutive telephone calls, or the interval between consecutive accidents to the same worker.

To test the hypothesis H' that the location parameter B is equal to some fixed value, it being assumed that the scale parameter σ is known, we can for simplicity take the set Ω of admissible populations from which the sample might have been drawn to be $\{-\infty < B < +\infty, \sigma = 1\}$, while the subset ω from which the sample must come when the hypothesis is true is $\{B = 0, \sigma = 1\}$. Then the likelihood-ratio λ_1 for testing this hypothesis is

$$\lambda_1 = \frac{P(\omega \text{ max.})}{P(\Omega \text{ max.})} = \frac{e^{-\sum_{i=1}^{n} x_i}}{e^{-\sum_{i=1}^{n} (x_i - x_1)}} = e^{-nx_1},$$

where x_1 is the smallest observation in a random sample of n. The region of acceptance of this hypothesis consists of all points in sample space for which

$$\lambda_{1\epsilon} \leq \lambda_1 \leq 1,$$

where $\lambda_{1\epsilon}$ is chosen so that $\int_{\lambda_{1\epsilon}}^{1} g_1(\lambda_1)\, d\lambda_1 = 1 - \alpha$, α being the level of significance used and $g(\lambda_1)\, d\lambda_1$ being the distribution of λ_1 when B is really equal to zero. The region $\lambda_{1\epsilon} \leq \lambda_1 \leq 1$ is equivalent to the region in the sample space for which

$$0 \leq x_1 \leq k_1 ; k_1 = -\frac{\log \lambda_{1\epsilon}}{n}.$$

For any value of B the distribution of x_1 is known [3] to be

$$\phi_1(x_1)\, dx_1 = n e^{-n(x_1 - B)}\, dx_1$$

Setting $B = 0$, the relationship between k_1 and α is

$$\int_0^{k_1} ne^{-nx_1}\,dx_1 = 1 - \alpha, \quad \text{so} \quad e^{-nk_1} = \alpha.$$

When $B \leq 0$, the power function $P(B)$, for this test is

$$P(B) = 1 - \int_0^{k_1} ne^{-n(x_1-B)}\,dx_1 = 1 - e^{nB}[1 - \alpha].$$

When $0 \leq B \leq k_1$, $P(B) = 1 - \int_B^{k_1} ne^{-n(x_1-B)}\,dx_1 = \alpha e^{nB}$. When $B \geq k_1$, $P(B) = 1$.

Since $e^{nB} > 1$ if $B > 0$ and also $e^{nB} < 1$ if $B < 0$, $P(B)$ is obviously $> \alpha$ if $B \neq 0$. This test is therefore completely unbiased in the sense of Daly [4]. In addition, it is not difficult to prove that this test has the unusual property of being a uniformly most powerful test with respect to all alternatives.

To test the hypothesis H'' that the location parameter is equal to some fixed value, say $B = 0$, when the scale parameter σ is unknown, the likelihood-ratio is easily seen to be

$$\lambda_2 = \left[\frac{\sum_{i=1}^{n}(x_i - x_1)}{\sum_{i=1}^{n} x_i}\right]^n = \left[\frac{1}{1 + \dfrac{nx_1}{\sum_{i=1}^{n}(x_i - x_1)}}\right]^n.$$

The region of acceptance consists of all points in the sample space for which $\lambda_{2\epsilon} \leq \lambda_2 \leq 1$ where $\int_{\lambda_{2\epsilon}}^{1} g_2(\lambda_2)\,d\lambda_2 = 1 - \alpha$. This is equivalent to the region

$$(1) \qquad 0 \leq \left[\frac{n(n-1)x_1}{\sum_{i=1}^{n}(x_i - x_1)} = t\right] \leq k_2\,; \qquad k_2 = (n-1)\frac{(1 - \lambda_{2\epsilon}^{1/n})}{\lambda_{2\epsilon}^{1/n}}.$$

The relation between k_2 and α is easily found from the distribution of t when $B = 0$, which is known to be [3]

$$\phi_2(t)\,dt = \frac{dt}{\left[1 + \dfrac{t}{n-1}\right]^n}.$$

Therefore $\int_0^{k_2} \phi_2(t)\,dt = 1 - \alpha$, so $\left[1 + \dfrac{k_2}{n-1}\right]^{-(n-1)} = \alpha$.

It is somewhat easier to find the power function of this test by considering the region of acceptance as made up of points in the x_1, s plane for which

$$0 \leq x_1 \leq \frac{k_2 s}{n} \quad \text{where} \quad s = \frac{\sum_{i=1}^{n}(x_i - x_1)}{n-1},$$

which is identical with the region in (1).

The joint distribution of x_1 and s is [3]

$$\psi_1(x_1, s)\, dx_1\, ds = \phi_3(x_1)\, dx_1 \cdot \phi_4(s)\, ds,$$

where

$$\phi_3(x_1)\, dx_1 = \frac{n}{\sigma} e^{-n(x_1 - B)/\sigma}\, dx_1$$

and

$$\phi_4(s)\, ds = \frac{\left(\frac{n-1}{\sigma}\right)^{n-1} s^{n-2} e^{-(n-1)s/\sigma}\, ds}{(n-2)!}.$$

When $B \leq 0$, the power function $P(B)$ of this test is

$$P(B) = 1 - \int_0^\infty ds \int_0^{k_2 s/n} \psi_1(x_1 s)\, dx_1 = 1 - e^{nB/\sigma}[1 - \alpha].$$

When $B \geq 0$, the power function is

$$\begin{aligned} P(B) &= 1 - \int_{Bn/k_2}^\infty ds \int_B^{k_2 s/n} \psi_1(x_1, s)\, dx_1 \\ &= \alpha e^{nB/\sigma} + I\left[n-1; \frac{n(n-1)B}{\sigma k_2}\right] - \alpha e^{nB/\sigma} I\left[n-1; \frac{n(n-1+k_2)B}{\sigma k_2}\right], \end{aligned} \tag{2}$$

where $I[p; x] = \dfrac{\Gamma_x(p)}{\Gamma(p)} = \dfrac{\int_0^x x^{p-1} e^{-x}\, dx}{\int_0^\infty x^{p-1} e^{-x}\, dx}$,

which is the form in which the Incomplete Gamma Function has been tabulated [5].

Since σ must be positive, $e^{nB/\sigma} < 1$ if $B < 0$ and therefore $P(B) > \alpha$ in the interval $-\infty < B < 0$. To show that $P(B)$ is $> \alpha$ in the interval $0 < B < \infty$, it is simpler to work with the expression for $P(B)$ as a double integral in (2), than to differentiate the power function directly. Performing the integration with respect to x_1,

$$P(B) = 1 + \int_{Bn/k_2}^\infty [e^{-(k_2 s - Bn)/\sigma} - 1] \cdot \phi_4(s)\, ds.$$

Differentiating with respect to B,

$$P'(B) = \int_{Bn/k_2}^\infty \frac{n}{\sigma} e^{-(k_2 s - Bn)/\sigma} \phi_4(s)\, ds.$$

The integral expression for $P'(B)$ is obviously positive. Therefore since for $B > 0$ the derivative is always positive the function must be monotomically

increasing in this interval ($0 < B < +\infty$), so $P(B)$ is $> \alpha$ when $B > 0$ Therefore this test is also completely unbiased.

We now consider the hypothesis H''' that two samples are drawn from exponential distributions with the same location parameter, assuming it is known the samples must have come from two exponential distributions with the same scale parameter. Given a sample of n_1 values of x drawn from $\frac{1}{\sigma} e^{-(x-B_1)/\sigma}\,dx$ and another independent sample of n_2 values of y drawn from $\frac{1}{\sigma} e^{-(y-B_2)/\sigma}\,dy$, the hypothesis we wish to test is that $B_2 = B_1$. Let x_1 be the smallest of the n_1 values of x and y_1 be the smallest of the n_2 values of y, let L be the smallest of the $n_1 + n_2 = N$ values of both x and y. Then the likelihood ratio for this hypothesis is

$$\lambda_3 = \left[\frac{\sum_{i=1}^{n_1}(x_i - x_1) + \sum_{i=1}^{n_2}(y_i - y_1)}{\sum_{i=1}^{n_1}(x_i - L) + \sum_{i=1}^{n_2}(y_i - L)}\right]^N = \left[\frac{1}{1 + \frac{z}{u}}\right]^N,$$

where
$$z = n_2(y_1 - x_1), \quad \text{if} \quad y_1 > x_1$$
$$= n_1(x_1 - y_1), \quad \text{if} \quad x_1 > y_1,$$

and
$$u = \sum_{i=1}^{n_1}(x_i - x_1) + \sum_{i=1}^{n_2}(y_i - y_1).$$

The region of acceptance, $\lambda_{3\alpha} \leq \lambda_3 \leq 1$, is equivalent to the region $0 \leq Z \leq K_3 u$, where K_3 is again a function of α, the level of significance, the exact relation being

$$\int_0^{k_3} \frac{(N-2)\,dt}{(1+t)^{N-1}} = 1 - \alpha, \qquad \text{so} \qquad \frac{1}{(1+k_3)^{N-2}} = \alpha.$$

It is known [3] that u is independent of Z, and that its distribution is

$$\phi_5(u)\,du = \frac{u^{N-3}e^{-u/\sigma}\,du}{\sigma^{N-2}(N-3)!}.$$

The distribution of z is somewhat complicated; but it can be derived by observing that the probability that z lies in any infinitesimal interval $z_1 \pm \frac{1}{2}\,dz_1$ is the sum of the probabilities that $n_2(y_1 - x_1)$ and $n_1(x_1 - y_1)$ lie in that interval and by then using standard methods for finding the distribution of the difference of two variates. For the case $G = B_2 - B_1 \geq 0$, the distribution $f(z)$ of z is

$$(3) \qquad \begin{aligned} f_1(z)\,dz &= \frac{e^{-n_1 G/\sigma}}{(n_1+n_2)\sigma}\left[n_1 e^{n_1 z/n_2\sigma} + n_2 e^{-z/\sigma}\right]dz, \qquad 0 \leq z \leq n_2 G, \\ f_2(z)\,dz &= \frac{\left[n_1 e^{n_2 G/\sigma} + n_2 e^{-n_1 G/\sigma}\right]e^{-z/\sigma}\,dz}{(n_1+n_2)\sigma}, \qquad n_2 G \leq z \leq \infty. \end{aligned}$$

For the case $G \leq 0$, the distribution of z can be derived from (3) by interchanging n_1 and n_2 and putting $-G$ in place of G.

The power function of this test can now be derived. For the case $G \geq 0$, the power function $P(G)$ is

$$P(G) = 1 - \left\{ \int_0^\infty du \int_0^{n_2 G} f_1(z)\phi_5(u)\, dz - \int_0^{n_2 G/k_3} du \int_{k_3 u}^{n_2 G} f_1(z)\phi_5(u)\, dz + \int_{n_2 G/k_3}^\infty du \int_{n_2 G}^{k_3 u} f_2(z)\phi_5(u)\, dz \right\}. \tag{4}$$

Upon integrating out and simplifying, the power function becomes

$$P(G) = \alpha\left(\frac{n_2 e^{-n_1 G/\sigma}}{n_1 + n_2}\right) + I\left[N-2; \frac{n_2 G}{k_3 \sigma}\right] + \alpha\left(\frac{n_1 e^{n_2 G/\sigma}}{n_1 + n_2}\right)\left\{1 - I\left[N-2, \frac{n_2 G(1+k_3)}{k_3\sigma}\right]\right\} - \frac{n_2}{n_1+n_2} e^{-n_1 G/\sigma}\left(\frac{n_2}{n_2 - n_1 k_3}\right)^{N-2} I\left[N-2; \frac{G(n_2 - n_1 k_3)}{k_3 \sigma}\right]$$

The power function when $G \leq 0$ is easily derived from that for $G \geq 0$ by everywhere interchanging n_1 and n_2 and substituting $-G$ for G.

To show that $P(G) > \alpha$ when $G \neq 0$, it is only necessary to show that the derivative $P'(G)$ of the power function is always positive when $G > 0$, and always negative when $G < 0$ It is again considerably simpler to use the expression for $P(G)$ as a double integral. For the case $G > 0$, integrating with respect to z in (4),

$$P(G) = 1 - \frac{n_2}{n_1+n_2}\left[1 - e^{-G(n_1+n_2)/\sigma}\right] + \int_0^{n_2 G/k_3} \frac{n_2 e^{-n_1 G/\sigma}}{n_1+n_2}\left[e^{n_1 z/n_2\sigma} - e^{-z/\sigma}\right]_{k_3 u}^{n_2 G} \phi_5(u)\, du - \int_{n_2 G/k_3}^\infty \frac{(n_1 e^{n_2 G/\sigma} + n_2 e^{-n_1 G/\sigma})}{n_1+n_2} \cdot \left[-e^{-z/\sigma}\right]_{n_2 G}^{k_3 u} \phi_5(u)\, du,$$

where $[f(x)]_a^b = f(b) - f(a)$. Upon differentiating and simplifying,

$$P'(G) = \frac{n_1 n_2}{(n_1+n_2)\sigma} \int_0^{n_2 G/k_3} e^{-n_1 G/\sigma}\left[e^{n_1 k_3 u/n_2\sigma} - e^{-k_3 u/\sigma}\right]\phi_5(u)\, du + \frac{n_1 n_2}{(n_1+n_2)\sigma} \int_{n_2 G/k_3}^\infty e^{-k_3 u/\sigma}\left[e^{n_2 G/\sigma} - e^{-n_1 G/\sigma}\right]\phi_5(u)\, du.$$

Both integrals are easily seen to always be positive, so $P'(G)$ is positive when $G > 0$. In the same manner it can be shown that $P'(G)$ is negative when $G < 0$. Therefore this test is also completely unbiased.

The question of investigating the bias of the likelihood-ratio tests for (a) testing the hypothesis that $\sigma = \sigma_0$ when B is known and (b) testing the hypothesis that $\sigma = \sigma_0$, nothing being known about the value of B, are practically identical with the analogous problems for a normal distribution. The results are also the same, for the λ test for (a) is completely unbiased, while that for (b) is biased.

REFERENCES

[1] J. NEYMAN and E. S. PEARSON, "On the use and interpretation of certain test criteria for purposes of statistical inference," *Biometrika*, Vol. 20a (1928), pp. 221–230.

[2] P. V. SUKHATME, "On the analysis of k samples from exponential populations with special reference to the problem of random intervals," *Stat. Res. Memoirs*, Vol. 1, pp. 94–112.

[3] P. V. SUKHATME, "Tests of significance for samples of the χ^2 population with 2 degrees of freedom," *Annals of Eugenics*, Vol. 8 (1937–38), pp. 54–55.

[4] JOSEPH F. DALY, "On the unbiased character of likelihood-ratio tests for independence in normal systems," *Annals of Math. Stat.*, Vol. 11 (1940), p. 2.

[5] K. PEARSON (Editor), *Tables of the Incomplete Gamma Function*, Biometric Laboratory, London, 1922.

ON THE MATHEMATICALLY SIGNIFICANT FIGURES IN THE SOLUTION OF SIMULTANEOUS LINEAR EQUATIONS

By L. B. Tuckerman

The National Bureau of Standards

1. **Introduction.** The number of mathematically significant figures in the solution of simultaneous linear equations has received attention from a number of writers [1–6]. It is an important subject, not only in least squares and correlations, but in many other problems of science where simultaneous equations arise: it may not be amiss, therefore, to examine it from a fresh start, particularly since (as will be shown) some of the rules that have been published on it fail in certain frequently occurring circumstances

2. **Definitions.** Before proceeding into the subject it will be necessary to distinguish between the computer's terms "significant figures" and "determinate significant figures" The former are the figures that compose a number, without the consecutive ciphers that precede or follow them, merely to locate the decimal point. "Determinate significant figures," on the other hand, are figures that are justifiable on computational grounds. From the computer's point of view, the number of significant figures remains independent of what is statistically significant. To avoid confusion in what follows, the term "significant figures" will be used in the computer's sense, and the adjective "determinate" will be supplied where mathematical determinacy is implied.

To avoid prolixity the term "observational error" will include any uncertainty arising either from errors in the observations or from the statistical nature of the problem (e g. sampling errors, grouping errors, etc). *The observational error of the result is independent of the particular sequence of computation followed and the accuracy with which it is carried out*

The term "computational error" will include all the additional uncertainties arising from the approximations occurring in the particular sequence of computation used, including the "rounding off" of the final result. *The computational errors, unlike the observational errors, depend in general upon the sequence of the intermediate steps used in the computation as well as on the number of significant figures to which they are carried.*

3. **Criterion of an adequate computation.** If the number written down at the end of a computation is to serve its purpose the maximum possible computational error must be suitably limited.

A decimal representation of a number containing f significant figures is subject

to an uncertainty (upper limit of absolute error) of 5 in the $(f + 1)$th place. It has, therefore, a possible relative (not absolute) error of representation somewhere between $5 \times 10^{-(f+1)}$ and 5×10^{-f}, in magnitude. This relative computational error sets the limit to any valid final rounding off. Regardless of the accuracy to which the intermediate steps of the computation have been carried, this relative computational error introduced by the final rounding off alone must be suitably limited.

In case all of the accuracy obtainable from the data is not needed in the result, the sum of the maximum possible computational error (including the error of the final rounding off) and the maximum possible observational error must be kept below the error which can be tolerated in the result.

In case all of the accuracy obtainable from the data is needed in the result, the maximum possible computational error in the result (including the error of the final rounding off) must be negligible in comparison with the uncertainty (observational error) in the result arising from uncertainty in the data. *Just how small a fraction of the observational error is "negligible" is necessarily a matter of judgment, and will depend upon the nature of the problem.* A computational error that would be wholly negligible in some ordinary computations might be intolerably large in the adjustment of an accurate geodetic survey. In any case the only basis for a valid judgment of the adequacy of the computation lies in a comparison of (i) the maximum possible computational error that can arise in the sequence of computations including the final "rounding off," with (ii) the observational error of the result arising from the observational errors inherent in the data.

4. **Propagation of error in a system of linear equations.** Assume that

$$\sum_t a_{st} x_t = b_s, \qquad s = 1, 2, \cdots, n, \tag{1}$$

is a set of simultaneous linear equations derived in some way from observations and in which the coefficients a_{st} and the absolute terms b_s may all be subject to observational error. If the relative (not absolute) observational error of a quantity q be represented by δ_q it may readily be seen that

$$\begin{cases} \delta x_j = -\sum_h \sum_k (x_k/x_j) A_{hj} a_{hk} \delta a_{hk} + \sum_s (b_s/x_j) A_{sj} \delta b_s \\ \delta \Delta = \sum_h \sum_k A_{hk} a_{hk} \delta a_{hk} \end{cases} \tag{2}$$

where Δ is the determinant of the coefficients a_{hk}, and A_{hk} is the term corresponding to a_{hk} in the reciprocal (not the adjoint) determinant.

5. **Upper limits to observational errors.** The sign and magnitude of the relative errors δa_{hk} and δb_s are unknown, but we shall assume that it is possible

in any problem to assign to them *upper limits*

$$|\,\delta a_{hk}\,| \quad \text{and} \quad |\,\delta b_s\,|$$

which in magnitude they cannot exceed. If the problem is such that the values of each of the δa_{hk} and the δb_s are wholly independent of each other, it is then possible that their magnitudes may all reach their upper limits $|\,\delta a_{hk}\,|$ and $|\,\delta b_s\,|$ simultaneously, in which case *upper bounds* of δx_j and $\delta\Delta$ may be placed at

$$(3)\qquad \begin{cases} |\,\delta x_j\,| = \sum_h \sum_k |\,(x_k/x_j)A_{hj}a_{hk}\,|\,|\,\delta a_{hk}\,| + \sum_s |\,(b_s/x_j)A_{sj}\,|\,|\,\delta b_s\,| \\ |\,\delta\Delta\,| = \sum_h \sum_k |\,A_{hk}a_{hk}\,|\,|\,\delta a_{hk}\,| \end{cases}$$

6. **Indefiniteness of the problem in the general case.** The values of the δa_{hk} and δb_s may not be independent of each other, in which circumstance knowledge of the law of their dependence would make it possible to assign upper limits to the magnitudes of δx_j and $\delta\Delta$ These upper limits can not be larger than the upper bounds shown in equation (3), and in special cases they will be much smaller. Since the dependence of δa_{hk} and δb_s may in general have any form whatever, cases can and will occur in which the upper limits of the relative errors of δx_j and $\delta\Delta$ may have any ratio whatever.

7. **Case of independent errors.** Any general discussion of the errors that can occur in x_j and Δ must be based either on some special assumption or on the limiting assumption that the errors are independent. It is this latter assumption that underlies the usual discussion, and will be the basis of what follows. Equation (3) gives the upper limit to the δx_j and $\delta\Delta$ under these assumptions.

8. **The ratios of $|\,\delta x_j\,|$ and $|\,\delta\Delta\,|$ are still indefinite in spite of the assumption of independent errors in the coefficients.** However, equation (3) does not determine any definite ratio or inequality between the upper bounds $|\,\delta x_j\,|$ and $|\,\delta\Delta\,|$. The nature of the observations may be such that some of the errors in the a_{hk} and b_s are very small and some relatively large. Not infrequently it is safe to assume that some of them are free from appreciable error and to ascribe all the error of the x_j to the error in one or two of the a_{hk} or b_s. If any statement of a definite relationship, either as an equality or an inequality between $|\,\delta\Delta\,|$ and the $|\,\delta x_j\,|$ is valid for all possible sets of linear equations, it must at least hold in the special case in which the errors of all the b_s and the errors of all except one of the a_{hk} are negligible.

If such a statement of a definite general relationship between these upper limits of errors can be made, it must be possible to write down an equation or an inequality between any one of the expressions $|\,A_{hk}\,|$ and some or all of the corresponding expressions $|\,(x_k/x_j)A_{hj}\,|$, $j = 1, 2, \cdots, n$, that will remain true no matter what be the values of the a_{hk} and the b_s in the original set of simultaneous equations It is obvious that the ratio of $|\,A_{hk}\,|$ and $|\,(x_k/x_j)A_{hj}\,|$, $(j \neq k)$, depends upon the values of the a_{hk}, and sets of equations can be found

to give any assigned value to that ratio. It is therefore impossible to state any rule that will restrict the ratio of the relative error of Δ and the relative error of any one of the x_j, valid for all possible sets of linear equations.

9. **Definite statement about the sum of the relative errors in the unknowns.** However, in the summation $\sum_j |\delta x_j|$ there occurs the term corresponding to $j = k$, for which $|(x_k/x_j)A_{hj}| = |A_{hk}|$, so that under the assumption that the a_{hk} and b_h are independent sources of error, we may write the inequality

$$\sum_j |\delta x_j| \not< |\delta\Delta| \tag{4}$$

which states that the sum of the upper bounds to the relative errors of all the x_j cannot be less than the upper bound to the relative error of the determinant Δ. A corresponding statement can easily be proved for the standard deviations

A limiting case can be constructed in which the inequality (4) reduces to

$$\sum_j |\delta x_j| = |\delta\Delta| \tag{5}$$

and in which all of the $|\delta x_j|$ are equal. For this case,

$$|\delta\Delta| = n\,|\delta x_j| \text{ for all values of } j. \tag{6}$$

If $n \not< 10$ it is obvious that there will be at least one more determinate significant figure in each of the x_j than in the determinant Δ of the coefficients.

It is frequently assumed that the number of determinate significant figures in the solution for any unknown cannot exceed the number of determinate significant figures in the determinant Δ of the coefficients. We see now that this statement can not be generally valid, even under the assumption that the a_{hk} and b_h are independent sources of error. As a matter of fact, it is necessary in some cases to compute some or even all of the unknowns to more significant figures than are determinate in the determinant Δ of the coefficients, if one would retain in the result all the accuracy that is obtainable from the data.

Cases in which the relative observational error of every one of the unknowns is less than the relative error of the determinant Δ probably occur rarely in practice; in fact the only ones that I have seen are those that I constructed purposely to show that such a thing is possible. However, cases in which the relative errors of one or several but not all of the unknowns are much smaller than the relative error of the determinant Δ, occur fairly frequently.

10. **Remarks on the case of "near indeterminacy."** The major interest in curve fitting centers around the condition of "near indeterminacy," i.e., of a small or near vanishing determinant Δ. Even in the circumstance where the relative error of the determinant is much greater than the relative error of some or all of the coefficients and absolute terms, the relative error of one or more of the unknowns may be much smaller than the relative error of the determinant, as may be seen from what follows.

In accurate experimentation the endeavor is, wherever possible, to arrange the experiment so that the quantity sought comes directly from the measurement as represented by an equation such as

$$x = p. \tag{7}$$

However, so ideal an experimental arrangement is rarely if ever possible, and it is a common experience to find that the measurements are represented by an equation such as

$$x + qy + rz + su + \cdots = p, \tag{8}$$

where qy, rz, su, etc., are small corrections that must somehow be evaluated. For simplicity, the discussion will be confined to the almost trivial case

$$x + qy = p. \tag{9}$$

Not infrequently the only way the correction can be evaluated is to rearrange the conditions of the experiment so that another equation is obtained in the form

$$x + q'y = p'. \tag{10}$$

Sometimes the nature of the experiment is such that it is not possible to change the coefficient of y by more than a small amount, under which conditions

$$q' = q(1 + \beta), \tag{11}$$

and

$$p' = p(1 + \alpha), \tag{12}$$

where β and α are small in comparison with 1. The solution of equations (9) and (10) now gives

$$x = \frac{\begin{vmatrix} p & q \\ p' & q' \end{vmatrix}}{\begin{vmatrix} 1 & q \\ 1 & q' \end{vmatrix}} = \frac{pq' - p'q}{q' - q} = p(1 - \alpha/\beta). \tag{13}$$

The quantity $q' - q$ seen in the denominator of this equation is the determinant Δ of the coefficients, and by equation (11) its value is βq. Since βq is assumed to be small here, the solution for x encounters a near vanishing denominator. It would, however, be wrong to assume that the number of determinate significant figures in x that can be obtained by solving the equations is necessarily limited to the number of determinate significant figures in the denominator Δ.

If the experimenter has been fortunate in finding suitable experimental conditions, the denominator $\Delta = \beta q$, although small in comparison with either q' or q, will still not cause difficulty. It will be observed that the coefficients of q' and q in the denominator are equal (both being unity). Now if the coefficients p and p' in the numerator are nearly enough equal, so that q' and q occur in both

numerator and denominator so nearly proportionally that the uncertainties in q and q' produce nearly compensating errors in both numerator and denominator, then x will be given to more determinate significant figures than are found in the denominator Δ. It can then be said that the experiment is successful in evaluating the correction term qy in equation (9).

On the other hand, in less fortunate circumstances, to the exasperation of the experimenter, the denominator $\Delta = q' - q = \beta q$ is not only small, but p' and p, although still nearly equal, differ enough so that the errors in q' and q are not compensated by the nearly equal coefficients in the numerator. The experiment will then fail to improve the approximation p for x by failing to evaluate the small correction qy in equation (9). This would be an inherent defect in the experiment and could not be removed by any manner of computation.

The same conclusion would of course be drawn from the coefficient of p (viz., $1 - \alpha/\beta$) at the extreme right of equation (13). It is not the size of β that alone determines the number of determinate significant figures in x, it is rather the ratio between α and β. In the fortunate experimental circumstances described above, the near equality of p' and p offsets the near equality of q' and q by reducing the term α/β to a value small compared with unity; the term α/β, being small, acts to reduce the effect of the uncertainties in q and q' (i.e., in q and β) in the evaluation of x. On the other hand, in less fortunate circumstances, the correction term α/β can not now shield x from the uncertainties in q and q' since the relative difference α between p and p' is not small enough to reduce α/β to innocuity.

11. **Numerical illustration of compensating errors.** As a "horrible example" especially constructed to emphasize the theoretical possibilities, take the following special case—

$$\begin{cases} 1000.10000x + 10.00000y = 1010.10000 \\ 1000.00000x + 10.00000y = 1010.00000 \end{cases} \tag{14}$$

wherein it is assumed that the coefficients and the absolute terms (assumed to be derived from the observational data) are all correct to the fifth decimal place as given, and no closer estimate of their errors is possible. So far as known, the upper limit to the absolute observational error of each is then the same, i.e. 5×10^{-6}, but the coefficients of x (a_{11} and a_{21}), and the absolute terms (b_1 and b_2), all have nine determinate significant figures, while the coefficients of y (a_{12} and a_{22}), have only seven. Thus,

$$|\delta a_{11}| \not> 5 \times 10^{-9}, \quad |\delta a_{21}| \not> 5 \times 10^{-9}, \quad |\delta b_1| \not> 5 \times 10^{-9}, \quad |\delta b_2| \not> 5 \times 10^{-9},$$

but

$$|\delta a_{12}| \not> 5 \times 10^{-7}, \quad |\delta a_{22}| \not> 5 \times 10^{-7}, \tag{15}$$

and $x = 1$, $y = 1$, $\Delta = 1$, whereupon a substitution of values from (15) into (3) gives the inequalities

$$(16) \quad |\delta x| \not> 3 \times 10^{-4}, \qquad |\delta y| \not> 3 \times 10^{-2}, \qquad |\delta \Delta| \not> 1.01 \times 10^{-2}.$$

So far as known, the determinant Δ may thus be in error by as much as 1 per cent, and y by as much as 3 per cent, yet x is known closer than 1/30th per cent. Here the value of the unknown x cannot be adequately represented by less than four significant figures, and might even require five, in spite of the fact that neither Δ nor y requires more than three significant figures to represent all that is certainly known about them.

The reason for this disparity in relative errors can be more easily seen by substituting numerical values for all the coefficients in the expression for x except a_{12} and a_{22}. The possible relative errors of a_{12} and a_{22} are, as noted above, about 100 times as great as the possible relative errors of a_{11}, a_{21}, b_1, and b_2, and are the controlling errors in Δ. In the solution

$$(17) \qquad x = \frac{1010.10000a_{22} - 1010.00000a_{12}}{1000.10000a_{22} - 1000.00000a_{12}},$$

however, both a_{12} and a_{22} occur in both numerator and denominator, and moreover the coefficient of each in the numerator is nearly equal to its coefficient in the denominator, so that a change in either a_{12} or a_{22} changes both numerator and denominator nearly proportionally, with the result that their ratio x is known much more accurately than either the numerator or the denominator Δ.

This kind of compensation of errors in a computation is not confined to the solution of simultaneous equations (and it is not an infrequent occurrence in other computations). This is one of the many reasons why it is impossible to give general rules for the retention of significant figures that will be valid for all types of computations

12. **Geometrical analogy.** Moulton [4] illustrated his reasoning by the following geometrical analogy. The solution of three linear equations is equivalent to finding the point of intersection of three planes. When the determinant of the coefficients is small in comparison with the coefficients themselves, these planes are either nearly parallel, or the line of intersection of any two of them is nearly parallel to the third In these cases small uncertainties in the location of any one of the planes correspond to large uncertainties in the position of their point of intersection.

In the first circumstance the planes might all be nearly parallel to one of the three coordinate planes, with the result that large uncertainty would afflict the value of the determinant and two of the unknowns, the third being much more accurately determined

In the second circumstance, the line of intersection of two of the planes might be nearly parallel to one of the coordinate axes. When that happens, large un-

certainty will afflict the value of the determinant, but only one of the unknowns, the other two being much more accurately determined.

This geometrical analogy can be extended to cover simultaneous equations with any number of unknowns. Near-vanishing of the determinant Δ of the coefficients necessarily implies relatively large uncertainties in the determinant and also in at least one of the unknowns, but not necessarily in all of them. These are, of course, very special cases, but, as noted above, they are of frequent occurrence in actual problems.

13. **Evaluation of computational error.** The relative computational error in x_i must be kept within certain definite limits which depend upon the particular problem to be solved (section 3). To do this it is necessary to be able to calculate an upper bound to the relative computational error inherent in any particular sequence of computations

In many computations it is easy to write down a simple formula that will set an upper bound to the relative computational error involved in that particular sequence This formula contains numbers f_1, f_2, f_3, etc., each representing the number of significant figures accurately computed at some particular step Once a simple formula for relative computational error is written down, it is easy to choose values of f_1, f_2, f_3, etc. that will give an upper bound to the relative computational error not larger than the permissible limit of maximum possible computational error outlined in section 3. This method of determining an upper bound of the relative computational error should be used whenever such a simple formula can be found. For example, to compute x from equation (13) we may use the following sequence: $r_1 = q' - q$, $r_2 = r_1/q = \beta$, $r_3 = p' - p$, $r_4 = r_3/p = \alpha$, $r_5 = r_4/r_2 = \alpha/\beta$, $r_6 = 1 - r_5 = 1 - \alpha/\beta$, $r_7 = pr_6 = p(1 - \alpha/\beta) = x$. x may then be written as a function of these partial results, viz.:

$$x = r_7 = pr_6 = p(1 - r_5) = p(1 - r_4/r_2) = p(1 - r_3/pr_2) = p(1 - r_4q/r_1). \tag{18}$$

Applying first order error theory we find

$$|\epsilon(x)| \leqq \left|\frac{\alpha/\beta}{1 - \alpha/\beta}\right| \{|\epsilon(r_1)| + |\epsilon(r_2)| + |\epsilon(r_3)| + |\epsilon(r_4)| + |\epsilon(r_5)|\} + |\epsilon(r_6)| + |\epsilon(r_7)| \tag{19}$$

where $\epsilon(r_i)$ represents the relative error in r_i arising from the computation by which r_i was determined from the preceding partial results, $r_1, r_2, \cdots, r_{i-1}$, and $\epsilon(x)$ is the total relative computational error in x when so computed. It is easy to keep $\epsilon(x)$ within any desired limits by suitably limiting each error term of (19). Since a computation accurate to f significant figures involves a relative computational error not greater than 5×10^{-f}, any desired limits can then be set to each error term of (19) by a proper choice of the number of significant figures that should be carried in that step.

Unfortunately there seem to be no reasonably simple formulae for determining upper bounds of the relative computational errors that arise in the solution of simultaneous linear equations in more than two variables This does not absolve the computer from the necessity of ensuring that his computational errors are suitably limited.

The method I have found most economical is to carry the solution of simultaneous linear equations to the capacity of the machine, and as each partial result r_i is obtained, write it as

$$r_i(1 \pm \epsilon_i),$$

where r_i is the value actually found and ϵ_i is a positive number representing the accumulation of uncertainty introduced by all preceding steps in the computation. At the end of the computation each of the unknowns is found in the form

$$x_j(1 \pm \epsilon_j), \tag{20}$$

where x_j represents the value found and ϵ_j is the upper bound of the relative computational error in x_j

A comparison of ϵ_j with the upper bound of the observational error $| \delta x_j |$ of equation (3) will then indicate whether the computation is adequate. If the comparison shows that the computation was inadequate, it will show in which steps the number of significant figures f_i was too small, and by how much. The computer can recompute, carrying these steps to the requisite number of figures with the assurance that his recomputation will then be adequate. The comparison will further indicate in which steps if any the number of significant figures f_i was larger than necessary.

When a computer has thus set suitable upper bounds to the relative computational error in the solution of a set of linear equations, he is in a position to plan solutions of future similar sets so as to perform his computations more economically and yet safely This is especially true when the solution of simultaneous linear equations arises week after week in routine testing.

14 **Conclusions.** Summary rules have been published, purporting to be safe guides to computers in avoiding needless work, and ensuring that the computations are carried to a sufficient degree of accuracy. Many of them are useful guides for certain types of computation and for limited ranges of the numerical values entering into the computation, but none of those that I have seen can be used generally. The only safe rule, where the matter is of importance, is to calculate the maximum possible computational error that can enter in the particular sequence of computation followed, and make sure that it is kept within the necessary limits.

It is sometimes necessary to carry the intermediate steps of a computation to many significant figures beyond the significant figures given in the data, or kept in the result The relative error of one of the unknowns may be very much smaller than the relative errors of the data from which it is computed, while the

relative error of another of the unknowns may be larger. The methods of ensuring that the computations are adequate are outlined in section 13

For the best sequence to follow in the elimination of the unknowns, I shall pass along a suggestion of Dr. W. Edwards Deming which he gave in one of our discussions of this subject. I venture to pass it along, because it has worked in every special case that I have constructed in an attempt to prove that it does not hold generally If ever the suggestion fails, the computer may change the sequence; but in any case he is obliged, as stated above, to calculate the maximum possible computational error that can enter into his calculations. Dr. Deming's suggestion is this: "To evaluate some but not all of the unknowns to the highest possible computational accuracy, retaining as few significant figures as possible in the intermediate steps, solve the equations by successive elimination, eliminating first and evaluating last the unknowns of greatest inherent relative accuracy."

15 **Summary.** Expressions are given for the maximum observational error in the unknowns of a system of simultaneous linear equations, in terms of the relative errors of the coefficients and absolute terms therein. In order to extract all the information possible from a system of linear equations representing observational results, it is not sufficient in general to assume that the relative errors in the unknowns are as large as the relative error in the determinant of the system. In many problems the computation of some of the unknowns must therefore be carried to more significant figures than are determinate in the determinant of the system. Methods are outlined for evaluating computational error in the solution of linear equations to ensure that the computations are adequate.

In conclusion I wish to express my thanks to Dr. W. Edwards Deming who has given much of his time to assist me in the preparation of this paper. He has made valuable suggestions on the material to be included and the general manner of presentation. In addition he has criticized the manuscript in detail and assisted in the final revision.

REFERENCES

[1] Edward B. Roessler, *Science*, Vol. 84, (1936), pp. 289–290.
[2] Joseph Berkson, ibid., p. 437.
[3] P. J. Rulon, ibid., pp. 483–484.
[4] F R. Moulton, ibid., pp. 574–575.
[5] W. Edwards Deming, *Science*, Vol. 85 (1937), pp. 451–454. *Least Squares* (Department of Agriculture Graduate School, 1938), pp. 105, 111, 121, 135.
[6] I. M. H. Etherington, *Edinb. Math Soc. Proc*, Vol. 3 (1932), Part 2, pp. 107–117

ON MECHANICAL TABULATION OF POLYNOMIALS

By J. C. McPherson

International Business Machines Corporation

1. Introduction. The purpose of this paper is to show how automatic accounting machines, which have been used previously in evaluating such quantities as Σx^n and $\Sigma x^{n-1}y$, may be used in the preparation of mathematical tables of integral powers, of polynomials, and of functions which can be approximated by polynomials. These tables may be prepared for any desired intervals of the argument such as 1, $\frac{1}{10}$, $\frac{1}{100}$, $\frac{1}{2}$, $\frac{1}{3}$, etc.

The method is an adaptation of the general theory of "cumulative" or "progressive" totals which has proved useful in computing moments and product moments both with and without accounting machines. The reader unfamiliar with the mathematical method and its machine applications might refer to such presentations as those of Hardy [1], Mendenhall and Warren [2, 3], Razram and Wagner [4], Brandt [5], and Dwyer [6, 7]. The main feature of the method is the computation of summed products or of summed powers by means of successive cumulated additions. It is shown in this paper how it is possible to use this same process in constructing tables of powers and tables of polynomials

2. The Cumulative Formulas. If the numbers F_x are defined and finite for $x = 1, 2, 3, \cdots, (a-1), a$, and if these values of F_x are cumulated for $x = a$, $x = a - 1$, etc, then the value in the row headed by $x = 1$ can be written as ${}^1T_1$. If these cumulations are cumulated successively with the superscript indicating the order of the cumulation and the subscript indicating the value of x which heads the row, then

$$ {}^2T_1 = \Sigma x F_x, \qquad {}^3T_1 = \Sigma \frac{(x+1)x}{2!} F_x, \qquad {}^3T_2 = \Sigma \frac{x(x-1)}{2!} F_x, $$

$$ {}^4T_1 = \Sigma \frac{(x+2)(x+1)x}{3!} F_x $$

and in general for $i < j$,

$$ {}^jT_i = \Sigma \frac{[x + j - (i+1)]^{(j-1)}}{(j-1)!} F_x. \tag{1} $$

Formula (1) is basic to much of the previous work involving cumulative totals. Various authors have studied such important special cases as (A) where F_x equals the frequency function f_x, (B) where $F_x = xf_x$, and (C) where F_x equals the sum of all the values of y having the same x value. These special cases have been found very useful in computing moments and product moments.

The moments may be expressed in terms of the cumulations in a variety of ways. The diagonal formulas have the differences of zero as coefficients and are expressed in terms of ${}^1T_1$, ${}^2T_1$, ${}^3T_2$, ${}^4T_3$, ${}^5T_4$, etc The columnar formulas, whose coefficients have been recently studied [6, 7], are expressed in terms of cumulations of the same order, ${}^jT_x$, with j fixed. Razram and Wagner [4] have given formulas which utilize the entries of different rows and different columns but which demand fewer entries for the formulas. Razram and Wagner worked out the formulas through $\Sigma x^4 f_x$ but the argument holds for $\Sigma x^4 F_x$. For purposes of comparison the values of $\Sigma x^i F_x$, $i = 0, 1, 2, 3, 4$, as they appear in the diagonal, columnar, and Razram-Wagner systems are presented in Table I.

TABLE I

Values of $\Sigma x^i F_x$ for $i = 0, 1, 2, 3, 4$.

F_x	Diagonal	Columnar	Razram-Wagner
ΣF_x	${}^1T_1$	${}^1T_1$	${}^1T_1$
$\Sigma x F_x$	${}^2T_1$	${}^2T_1$	${}^2T_1$
$\Sigma x^2 F_x$	${}^2T_1 + 2{}^3T_2$	${}^3T_1 + {}^3T_2$	${}^3T_1 + {}^3T_2 = {}^3T_{1+2}$
$\Sigma x^3 F_x$	${}^2T_1 + 6{}^3T_2 + 6{}^4T_3$	${}^4T_1 + 4{}^4T_2 + {}^4T_3$	${}^2T_1 + 6{}^4T_2$
$\Sigma x^4 F_x$	${}^2T_1 + 14{}^3T_2 + 36{}^4T_3 + 24{}^5T_4$	${}^5T_1 + 11{}^5T_2 + 11{}^5T_3 + {}^5T_4$	${}^3T_{1+2} + 12{}^5T_{2+3}$

In developing the theory of the later sections of this paper I have developed further formulas of the type shown by Razram and Wagner since these formulas have fewer terms than do those of the other systems and the coefficients are factorable by $(j-1)!/2$. These formulas for $\Sigma x^s F_x$, with s even, feature such terms as ${}^3T_1 + {}^3T_2 = {}^3T_{1+2}$, ${}^5T_{2+3}$, etc., so that there are two entries from the same column. For the purposes of this paper it is preferable to have a single entry from each column and this situation results from continued application of the formula

$$(2) \qquad {}^jT_{i+(i+1)} = {}^jT_i + {}^jT_{i+1} = {}^{j-1}T_i + 2\,{}^jT_{i+1}.$$

The formulas for $\Sigma x^s F_x$ with $s \leqq 12$ are given. The alternative forms are given for the formulas involving even values of s.

$$\Sigma F_x = {}^1T_1, \qquad \Sigma x F_x = {}^2T_1, \qquad \Sigma x\,{}^2F_x = {}^3T_1 + {}^3T_2 = {}^3T_{1+2} = {}^2T_1 + 2\,{}^3T_2,$$

$$\Sigma x\,{}^3F_x = {}^2T_1 + 6\,{}^4T_2, \qquad \Sigma x\,{}^4F_x = {}^3T_{1+2} + 12\,{}^5T_{2+3}$$

$$= {}^2T_1 + 2\,{}^3T_2 + 12\,{}^4T_2 + 24\,{}^5T_3,$$

$$\Sigma x\,{}^5F_x = {}^2T_1 + 30\,{}^4T_2 + 120\,{}^6T_3,$$

$$\Sigma x\,{}^6F_x = {}^3T_{1+2} + 60\,{}^5T_{2+3} + 360\,{}^7T_{3+4}$$

$$= {}^2T_1 + 2\,{}^3T_2 + 60\,{}^4T_2 + 120\,{}^5T_3 + 360\,{}^6T_3 + 720\,{}^7T_4,$$

$$\Sigma x\,{}^7F_x = {}^2T_1 + 126\,{}^4T_2 + 1680\,{}^6T_3 + 5040\,{}^8T_4,$$

$$\Sigma x\,{}^8F_x = {}^3T_{1+2} + 252\,{}^5T_{2+3} + 5040\,{}^7T_{3+4} + 20160\,{}^9T_{4+5}$$

$$= {}^2T_1 + 2\,{}^3T_2 + 252\,{}^4T_2 + 504\,{}^5T_3 + 5040\,{}^6T_3 + 10080\,{}^7T_4$$

$$+ 20160\,{}^8T_4 + 40320\,{}^9T_5,$$

$$(3)\quad \Sigma x\,{}^{9}F_x = {}^{2}T_1 + 510\,{}^{4}T_2 + 17640\,{}^{6}T_3 + 151200\,{}^{8}T_4 + 362880\,{}^{10}T_5,$$

$$\Sigma x\,{}^{10}F_x = {}^{3}T_{1+2} + 1020\,{}^{5}T_{2+3} + 52920\,{}^{7}T_{3+4} + 604800\,{}^{9}T_{4+5} + 1814400\,{}^{11}T_{5+6}$$

$$= {}^{2}T_1 + 2\,{}^{3}T_2 + 1020\,{}^{4}T_2 + 2040\,{}^{5}T_3 + 52920\,{}^{6}T_3 + 105840\,{}^{7}T_4$$

$$+ 604800\,{}^{8}T_4 + 1209600\,{}^{9}T_5 + 1814400\,{}^{10}T_6 + 3628800\,{}^{11}T_6,$$

$$\Sigma x\,{}^{11}F_x = {}^{2}T_1 + 2046\,{}^{4}T_2 + 168960\,{}^{6}T_3 + 3160080\,{}^{8}T_4 + 19958400\,{}^{10}T_5$$

$$+ 39916800\,{}^{12}T_6,$$

$$\Sigma x\,{}^{12}F_x = {}^{3}T_{1+2} + 4092\,{}^{5}T_{2+3} + 506880\,{}^{7}T_{3+4} + 12640320\,{}^{9}T_{4+5}$$

$$+ 99792000\,{}^{11}T_{5+6} + 239500800\,{}^{13}T_{6+7}$$

$$= {}^{2}T_1 + 2\,{}^{3}T_2 + 4092\,{}^{4}T_2 + 8184\,{}^{5}T_3 + 506880\,{}^{6}T_3 + 1013760\,{}^{7}T_4$$

$$+ 12640320\,{}^{8}T_4 + 25280640\,{}^{9}T_5 + 99792000\,{}^{10}T_5$$

$$+ 199584000\,{}^{11}T_6 + 239500800\,{}^{12}T_6 + 479001600\,{}^{13}T_7.$$

The derivation of these formulas is obtained with the use of (1), with the use of

$$(4)\qquad {}^{j}T_i = {}^{j}T_{i+1} + {}^{j-1}T_i,$$

and with the use of formulas of lower order For example we have from (1)

$$\Sigma \frac{(x+4)(x+3)(x+2)(x+1)x}{120} F_x = {}^{6}T_1$$

so that

$$\Sigma x\,{}^{5}F_x = 120\,{}^{6}T_1 - 10\,\Sigma x\,{}^{4}F_x - 35\,\Sigma x\,{}^{3}F_x - 50\,\Sigma x\,{}^{2}F_x - 24\,\Sigma x F_x$$

which after substitution of $\Sigma x\,{}^{4}F_x$, $\Sigma x\,{}^{3}F_x$, etc. and simplification results in the value ${}^{2}T_1 + 30\,{}^{4}T_2 + 120\,{}^{6}T_3$.

3. Tables of powers. If $F_x = 1$ when $x = a$, but is zero otherwise then $\Sigma x\,{}^{s}F_x$ is equal to a^s. It follows that the value of a^s can be obtained from the successive cumulations of this F_x with the use of (3) For example in Table II

TABLE II

Cumulations of $F_x = 1$, *when* $x = 6$,
0, *when* $x \neq 6$.

a	x	F_x	${}^{1}T$	${}^{2}T$	${}^{3}T$	${}^{4}T$	${}^{5}T$
1	6	1	1	1	1	1	1
2	5	0	1	2	3	4	5
3	4	0	1	3	6	10	15
4	3	0	1	4	10	20	*35*
5	2	0	1	5	*15*	*35*	70
6	1	0	1	*6*	21	56	126
7		0	1	7	28	84	210
8		0	1	8	36	120	330

$$6^2 = {}^2T_1 + 2\,{}^3T_2 = 6 + 2(15) = 36,$$

$$6^3 = {}^2T_1 + 6\,{}^4T_2 = 6 + 6(35) = 216,$$

$$6^4 = {}^2T_1 + 2\,{}^3T_2 + 12\,{}^4T_2 + 24\,{}^5T_3 = 6 + 2(15) + 12(35) + 24(35) = 1296.$$

The values of ${}^2T_1$, ${}^3T_2$, ${}^4T_2$ and ${}^5T_3$ for $a = 6$ are italicized in Table II

To get the values of 5^2, 5^3, 5^4, etc it would be necessary to start to cumulate from $x = 5$. Now since the values of ${}^1T_1$ are unity, it follows that the values for $a = 5$ can be found by taking the entries above those for $a = 6$ Thus ${}^2T_1 = 5$, ${}^3T_2 = 10$, ${}^4T_2 = 20$, ${}^5T_3 = 15$ with $5^2 = 5 + 2(10)$, $5^3 = 5 + 6(20)$, $5^4 = 5 + 2(10) + 12(20) + 24(15)$. It is evident in general that the values for any a^2, a^3, a^4 can be obtained by taking the row headed by a as the bottom row. Thus using $a = 8$, we have $8^2 = 8 + 2(28)$, $8^3 = 8 + 6(84)$, etc. It then appears that we may omit the x column of Table II and consider the cumulations to be ascending cumulations for a rather than descending cumulations for x.

A more satisfactory course is to cumulate the coefficients so as to eliminate the multiplications. Thus the value of ${}^6T_1$ could be obtained without multiplication by cumulating 6, 0, 0, 0, 0 $\cdots$ rather than 1, 0, 0, 0, $\cdots$. Several cumulations may be carried on at the same time so that the additions are not necessary and the tabulation results in a table of the desired powers.

In preparation of a power table, the formulas (3) become a series of instructions on the way in which we are to do the cumulating. For instance the formula:

$$x^7 = 5040\,{}^8T_4 + 1680\,{}^6T_3 + 126\,{}^4T_2 + {}^2T_1,$$

tells us that to form a table of the seventh power we must cumulate[1] the coefficient 5040 eight times; add in the coefficient 1680 when there are six operations; the coefficient 126 when there are four; and the coefficient 1 when there are two remaining. A change in subscript tells us that the coefficient when first included forms a separate total ahead of the ones already partly figured. When the subscript does not change, the coefficient is to be included in the first summary card total. The final cumulating operation prints the actual table.

To prepare a power table by machine we secure a set of cards punched all alike with the numbers from 1 to 9 punched diagonally in successive columns across the card. The machine is wired to add the coefficient of the highest term by selecting the proper digits from the diagonals, cumulate after each card and summary punch each total. This way of starting saves one cumulation. The summary cards are cumulated repeatedly in the same manner until the number of operations indicated by the highest term is completed. When the number of operations remaining equals j of another term ${}^jT_i$, a card for the coefficient of that term is included in the tabulation ahead of the summary cards. This automatically adds the new coefficient to each term of the series. When the subscript i in ${}^jT_i$ changes, the new coefficient card must form a separate total;

[1] This operation is generally known as *progressive totalling* in machine operation.

when it does not change, the coefficient card must tabulate in the first summary card total.

To illustrate the tabulation of power tables, the formula for the cube table is—

$x^3 = 6\,{}^4T_2 + {}^2T_1$

The successive operations yield the following table:

TABLE III

x	Operation number			
	1	2	3	$4.x^3$
1	0	0	1	1
2	6	6	7	8
3	6	12	19	27
4	6	18	37	64
5	6	24	61	125
6	6	30	91	216
7	6	36	127	343
8	6	42	169	512
9	6	48	217	729
10	6	54	271	1000

In actual machine work, operation 1 can be omitted and work begun with operation 2. The machine is set to add the coefficient 6 of the highest term from each card and an accumulated total is printed and punched for each card tabulated, giving the results shown under operation 2. An additional card is punched for the coefficient of the second term, 1, and placed ahead of the cards produced in operation 2. The cumulation and punching is repeated, giving the results shown under operation 3. The summary cards from this operation are cumulatively tabulated, giving the results shown under operation 4, which is the table of cubes desired.

Similarly, for a table of the fourth power, the formula $x^4 = 24\,{}^5T_3 + 12\,{}^4T_2 + 2\,{}^3T_2 + {}^2T_1$ indicates the following operations—

TABLE IV

x	Operation number				
	1	2	3	4	$5\,x^4$
1	0	0	0	1	1
2	0	12	14	15	16
3	24	36	50	65	81
4	24	60	110	175	256
5	24	84	194	369	625
6	24	108	302	671	1296
7	24	132	434	1105	2401
8	24	156	590	1695	4096
9	24	180	770	2465	6561
10	24	204	974	3439	10000
11	24	228	1202	4641	14641
12	24	252	1454	6095	20736

Note in operation 3 where the subscript does not change, the coefficient 2 is added to the first card punched by the machine, while in operation 4 where it changes, the coefficient 1 appears as a separate total.

4. Tables of polynomials. To tabulate values of $f(x) = a + bx + cx^2 \cdots$ (where $a, b, c, \cdots$, are positive or negative coefficients) the method is similar to that of preparing power tables except that the coefficients to be added are determined by multiplying the coefficients of the formulas for the different powers by the values a, b, c etc., adding the coefficients of like terms in the various formulas, and using these resultant coefficients in place of the simple coefficients used in the power tables. Thus if we wish to tabulate values of $f(x) = 4 + 3x + 2x^2 + x^5$ the coefficients are found as follows:

$$
\begin{aligned}
4x^0 &= 4\,{}^1T_0 \\
+\,3x &= \quad +\,3\,{}^2T_1 \\
+\,2x^2 &= \quad +\,2\,{}^2T_1 + 2\cdot 2\,{}^3T_2 \\
+\,x^5 &= \quad +\,{}^2T_1 \qquad\qquad +\,30\,{}^4T_2 + 120\,{}^6T_3 \\
\hline
f(x) &= 4\,{}^1T_0 + 6\,{}^2T_1 + \quad 4\,{}^3T_2 + 30\,{}^4T_2 + 120\,{}^6T_3
\end{aligned}
$$

This equation gives instructions to perform six operations with 120 as coefficient; adding the coefficient 30 as a separate total when there are 4 operations remaining; adding 4 to the first summary card total when there are 3 operations; adding 6 as a separate total when there are 2 operations remaining; and adding 4 on the last operation.

The first few totals appear thus—

TABLE V

x	*Operation number*					
	1	2	3	4	5	6: $f(x)$
0						4
1	0	0	0	0	6	10
2	0	0	30	34	40	50
3	120	120	150	184	224	274
4	120	240	390	574	798	1072
5	120	360	750	1324	2122	3194
6	120	480	1230	2554	4676	7870
7	120	600	1830	4384	9060	*16930*
8	120	720	2550	6934	*15994*	32924
9	120	840	*3390*	*10324*	26318	59242
10	*120*	*960*	4350	14674	40992	100234

It is not necessary to confine these tables to values for whole numbers, as we can tabulate equally well values of $f(x)$ for intervals of x of .1, 01 or .001 or $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$ etc. In this case, before combining formulas for different powers we multi-

ply both sides by the desired interval raised to the power to which x is raised in that particular formula, then add like terms as before.

To tabulate the previous example in $.1x$ intervals we proceed as follows:

$$
\begin{aligned}
4x^0 &= 4.000\ {}^1T_0 \\
3x/10 &= +\ .3\ {}^2T_1 \\
2(x/10)^2 &= +\ .02\ {}^2T_1 + .04\ {}^3T_2 \\
(x/10)^5 &= +\ .00001\ {}^2T_1 + .00030\ {}^4T_2 + .00120\ {}^6T_3 \\
\hline
f(x) &= 4\ {}^1T_0 + .32001\ {}^2T_1 + .04\ {}^3T_2 + .00030\ {}^4T_2 + .00120\ {}^6T_3
\end{aligned}
$$

TABLE VI

x	*Operation number*					
	1	2	3	4	5	6:$f(x)$
1	0	0	0	0	32001	4.32001
2	0	0	0003	0403	.36031	4 68032
3	0012	0012	0015	0418	.40211	5 08243
4	0012	0024	.0039	0457	44781	5.53024
5	.0012	0036	.0075	0532	50101	6 03125
6	0012	0048	0123	.0655	.56651	6.59776
7	0012	0060	0183	.0738	.64031	7 23807
8	.0012	0072	0255	.0903	73961	8.07768
9	0012	0084	0339	1332	87281	8.95049
10	0012	.0096	0435	1767	1.04951	10.00000

Where any coefficients are negative in the equations expressed in ${}^jT_i$ terms, they are simply added in as minus figures.

To round off the preceding function to 3 decimal places, we add 5 to the constant term ${}^1T_0$ in the position to the right of the last decimal retained, i.e. in this case the 4th decimal place. The constant term is then 4.0005.

Exact	*Counter reads*	*Prints*
4.32001	4.32051	4.320
4.68032	4.68082	4.680
5.08243	5.08293	5 082
5.53024	5.53074	5.530
6.03125	6 03175	6.031
6.59776	6 59826	6.598
7.23807	7.23857	7.238
8.07768	8.07818	8.078
8 95049	8.95099	8.950
10.00000	10.00050	10.000

5. Automatic calculation of polynomial coefficients. Frequently when polynomials are being evaluated, the process of forming the coefficients can be

performed automatically from a punched-card table. Such a table consists of a set of cards for each power x^s containing the multiples of all the coefficients of each of the terms ${}^jT_i$ in the formula (3) for that power. These multiples are 1, 2, 3, 4, $\cdots$, 9; 10, 20, 30, 40, $\cdots$, 90; 100, 200, $\cdots$, 900; 1000, 2000 etc., and may be produced automatically by making a linear table of each coefficient in the manner described in this paper. Each card is punched with the information called for by the heading of the following card form:

s	j	i	multiple	coeff. $\times$ multiple
07	06	03	00005	008400

The particular figures indicated are those which would be punched for the term $5(1680){}^6T_3$ in the representation of $5x^7$ according to formula (3).

The table is used by withdrawing the cards for the coefficients a, b, c, d, etc of the desired polynomial. For instance, if one of the polynomial coefficients is $14485\,x^7$, we select from the x^7 section of the table all cards containing the multiples 10000, 4000, 400, 80, and 5. In the x^7 table there are 4 cards for each multiple, one each for terms ${}^8T_4$, ${}^6T_3$, ${}^4T_2$, and ${}^2T_1$. These cards are combined with the cards selected for the other coefficients of the polynomial and sorted to bring all cards for each ${}^jT_i$ together. The cards for each term ${}^jT_i$ are then automatically added on the electric accounting machine.

6. Subdividing tables. In preparing tables it may be desired to prepare the table in more detail at certain points, giving values of the function at 1/10, 1/20, 1/50, or 1/100, etc., of the interval of the rest of the table. This may readily be done by recalculating the coefficients of the cumulative terms, and using these values in the same manner as the original ones.

There are many formulas for the determination of the subdivided differences given in various texts on interpolation, such as those given by Comrie [8] and Bower [9]. One effective method is to use formulas (3) to calculate the subdivided differences. The values called for in the formula for the highest power are taken from the table of the function at the regular interval, giving effect to the rule involving subscripts. These coefficients are reduced by an amount sufficient to cancel the coefficient of the highest cumulative term, and the coefficients of the remaining cumulative terms are reduced in proportion according to formula (3) for the highest power. Usually the coefficient of the highest term of the formula will divide evenly into the coefficient taken from the table, and the other reductions are calculated by multiplying this result by the other coefficients of the formula. The highest remaining coefficient is then reduced by an amount sufficient to cancel itself, and, by use of the formula (3) for the power whose highest cumulative term matches the highest remaining coefficient, the reduction to the remaining cumulative terms is calculated and subtracted.

The highest remaining coefficient is reduced in a like manner, and this process is continued until all the cumulative coefficients have been analyzed

The partial cumulative coefficients thus computed are multiplied by the desired subdivision $1/m$ raised to the power of the corresponding formula (3), and recombined to form the new coefficients, as shown in the example below. In taking values from the table, when the subscript does not change, the tabular value must be reduced by the amount of the higher coefficient with the same subscript, to give effect to the rule that the coefficients in such cases are increments (see last example in section 3).

To subdivide the polynomial of section 4 at $x = 7.0$, we take the italicized values from Table V starting at $f(7)$ as ${}^1T_0$, and proceed as follows:

	${}^6T_3$	${}^5T_3$	${}^4T_2$	${}^3T_2$	${}^2T_1$	${}^1T_0$
		960		10324		
From Table V . .	120	−120	3390	−3390	15994	16930
$F(x)$ . . .	120	840	3390	6934	15994	16930
ax^5	120		30		1	
		840	3360	6934	15993	
bx^4		840	420	70	35	
			2940	6864	15958	
cx^3			2940		490	
				6864	15468	
dx^2				6864	3432	
					12036	
ex.					12036	
						16930

If the interval is 1/10 we have:

	${}^6T_3$	${}^5T_3$	${}^4T_2$	${}^3T_2$	${}^2T_1$	${}^1T_0$
$x^5/10^5$ =	.00120		.00030		.00001	
$35x^4/10^4$ =		.0840	.04200	0070	.00350	
$490x^3/10^3$ =			2.94000		.49000	
$3432x^2/10^2$ =				68.6400	34.32000	
$12036x/10$ =					1203.60000	+16930

$f(x) = .00120\,{}^6T_3 + .0840\,{}^5T_3 + 2.9823\,{}^4T_2 + 68.6470\,{}^3T_2 + 1238.41351\,{}^2T_1 + 16930\,{}^1T_0$ provides the coefficients for subtabulating the function at the desired interval, beginning at the argument $x = 7.0$

7. Accuracy of Tables. When the values of the coefficients are not exact, owing to the original values for a, b, c etc. or the dropping of decimals in the computation of the coefficients, the errors accumulate fairly rapidly. Each coefficient will introduce its own error into the summation.

To maintain accuracy throughout a long table it is advisable to transform $f(x)$ by Horner's method of decreasing the roots [10, pp 100–101], compute new coefficients for the transformed equation at intervals, and prepare the table in sections. Decreasing the roots by r gives us a new starting point at $x = r$.

Since two or more functions may be computed at one time, a function for which the coefficients are not exact may be computed by adding in the usual way from the starting values and subtracting from the ending values simultaneously. As many digits as agree in both tabulations of the function may be considered correct.

The tabulations can be made to practically any degree of accuracy on the equipment available, as the newer machines can be formed into counters of any capacity up to 80 digits. In practice, counters of 16, 20 or 24 digits will ordinarily suffice for the accuracy desired and two or more functions can be evaluated simultaneously. Cards are read and added at the rate of 150 per minute, or read, added and listed on the tape at the rate of 80 per minute and new summary cards produced at the rate of 40 per minute (on alphabetic equipment with gang summary punches). Computation may be carried out with additional decimal places and the final tabulation of the function rounded off to the nearest number retained.

8. Summary. The cumulative or progressive-total method is shown to be applicable to the preparation of tables of functions expressed in the form of a power series.

The cumulative formulas for the powers through the twelfth power have been presented, and simple methods are given for transforming a power series into its corresponding cumulative formula, for changing the interval of the table, rounding off the values of the function, and subdividing the table at desired points.

It is hoped that this discussion will make tables in printed or punched-card form more generally available as a tool for the computer. Since tables may be so readily prepared by this process, the usefulness of the tabular method of solving problems is greatly increased.

The author wishes to acknowledge his thanks to Professor P. S. Dwyer for various suggestions, particularly in connection with section 2.

REFERENCES

[1] G. F Hardy, *Theory of Construction of Tables of Mortality*, pp. 59–62, 124–128.

[2] R. M. Mendenhall and R. Warren, *The Mendenhall-Warren-Hollerith Correlation Method*, Columbia University Statistical Bureau Document, No. 1, 1929, Columbia University, New York.

[3] R M Mendenhall and R. Warren, "Computing statistical coefficients from punched cards," *Jour. Educ. Psych.*, Vol 21 (1930), pp. 53–62.

[4] H. S. Razram and M. E Wagner, "The summation method in statistics," *Jour. Exper. Psych.*, Vol. 14 (1931), pp. 270–283.

[5] A E. Brandt, *Uses of the Progressive Digit Method*, Punched Card Method in Colleges and Universities, pp 423–436.

[6] P S. DWYER, "The computation of moments with the use of cumulative totals," *Annals of Math. Stat* , Vol. 9 (1938), pp. 288–304.
[7] P. S DWYER, "The cumulative numbers and their polynomials," *Annals of Math. Stat.*, Vol. 9 (1940), pp 66–71
[8] L. J. COMRIE, *Interpolation and Allied Tables*, Reprinted from the Nautical Almanac for 1937, His Majesty's Stationery Office.
[9] E C. BOWER, "Systematic subdivision of tables," Lick Observatory Bulletin 467 (1935), University of California Press.
[10] WHITTAKER and ROBINSON, *The Calculus of Observations*, Blackie and Son, Ltd., London, 1932.

ON THE PROBABILITY OF THE OCCURRENCE OF AT LEAST m EVENTS AMONG n ARBITRARY EVENTS

By Kai Lai Chung

Tsing Hua University, Kunming, China

Introduction. Let $E_1, \cdots, E_n$, denote n arbitrary events. Let $p_{\nu_1 \cdots \nu_i \nu_{i+1} \cdots \nu_j}$, where $0 \leqq i \leqq j \leqq n$ and $(\nu_1, \cdots, \nu_j)$ is a combination of the integers $(1, \cdots, n)$, denote the probability of the non-occurrence of $E_{\nu_1}, \cdots, E_{\nu_i}$ and the occurrence of $E_{\nu_{i+1}}, \cdots, E_{\nu_j}$. Let $p_{(\nu_1 \cdots \nu_i)}$ denote the probability of the occurrence of $E_{\nu_1}, \cdots, E_{\nu_i}$ and no others among the n events. Let $S_j = \Sigma p_{\nu_1 \cdots \nu_j}$ where the summation extends to all combinations of j of the n integers $(1, \cdots, n)$. Let $p_m(\nu_1, \cdots, \nu_k)$, $(1 \leqq m \leqq k \leqq n)$, denote the probability of the occurrence of at least m events among the k events $E_{\nu_1}, \cdots, E_{\nu_k}$.

By the set $(x_1, \cdots, x_b, \cdots, x_a) - (x_1, \cdots, x_b)$ (where $b \leqq a$) we mean the set $(x_{b+1}, \cdots, x_a)$. And by a $\binom{a}{b}$-combination out of $(x_1, \cdots, x_a)$ we mean a combination of b integers out of the a integers $(x_1, \cdots, x_a)$.

We often use summation signs with their meaning understood, thus for a fixed k, $1 \leqq k \leqq n$, the summations in $\Sigma p_{\nu_1 \cdots \nu_k}$, or $\Sigma p_m(\nu_1, \cdots, \nu_k)$, extend to all the $\binom{n}{k}$-combinations out of $(1, \cdots, n)$.

The following conventions concerning the binomial coefficients are made:

$$\binom{0}{0} = 1, \qquad \binom{a}{b} = 0 \qquad \text{if} \qquad a < b \qquad \text{or if} \qquad b < 0.$$

It is a fundamental theorem in the theory of probability that, if $E_1, \cdots, E_n$ are incompatible (or "mutually exclusive"), then

$$p_1(1, \cdots, n) = p_1 + \cdots + p_n .$$

When the events are arbitrary, we have Boole's inequality

$$p_1(1, \cdots, n) \leqq p_1 + \cdots + p_n .$$

Gumbel[1] has generalized this inequality to the following:

$$p_1(1, \cdots, n) \leqq \frac{\Sigma p_1(\nu_1, \cdots, \nu_k)}{\binom{n-1}{k-1}},$$

[1] *C. R. Acad. Sc.* Vol. 205(1937), p. 774.

for $k = 1, \cdots, n$. The case $k = 1$ gives Boole's inequality. Fréchet[2] has announced that Gumbel's result can be sharpened to the following

$$A_{k+1} = \frac{\Sigma p_1(\nu_1, \cdots, \nu_{k+1})}{\binom{n-1}{k}} \leqq \frac{\Sigma p_1(\nu_1, \cdots, \nu_k)}{\binom{n-1}{k-1}} = A_k, \tag{1}$$

for $k = 1, \cdots, n-1$. Thus, A_k is non-increasing for k increasing. On the other hand, Poincaré has obtained the following formula which expresses $p_1(1, \cdots, n)$ in terms of the S_i's,

$$\begin{aligned} p_1(1, \cdots, n) &= \sum p_{\nu_1} - \sum p_{\nu_1\nu_2} + \sum p_{\nu_1\nu_2\nu_3} - \cdots \\ &\quad + (-1)^n p_{1\cdots n} = \sum_{j=1}^{n} (-1)^{j-1} S_j. \end{aligned} \tag{2}$$

In the present paper we shall study the more general function $p_m(\nu_1, \cdots, \nu_k)$ as defined above. First we generalize Poincaré's formula and Fréchet's inequalities In Theorem 1 we establish (for $1 \leqq m \leqq n$)

$$\begin{aligned} p_m(1, \cdots, n) &= \sum p_{\nu_1\cdots\nu_m} - \binom{m}{1} \sum p_{\nu_1\cdots\nu_{m+1}} \\ &\quad + \binom{m+1}{2} \sum p_{\nu_1\cdots\nu_{m+2}} + \cdots + (-1)^{n-m} \binom{n-1}{m-1} p_{1\cdots n} \\ &= \sum_{i=0}^{n-m} (-1)^i \binom{m+i-1}{i} S_{m+i}. \end{aligned} \tag{3}$$

Although this result is well known, we prove it in preparation for Theorem 2. Theorem 3 establishes

$$A_{k+1}^{(m)} = \frac{\Sigma p_m(\nu_1, \cdots, \nu_{k+1})}{\binom{n-m}{k+1-m}} \leqq \frac{\Sigma p_m(\nu_1, \cdots, \nu_k)}{\binom{n-m}{k-m}} = A_k^{(m)}, \tag{4}$$

for $k = 1, \cdots, n-1$ and $1 \leqq m \leqq k$.

Next, we extend the inequalities (4), and in Theorem 4 we show that

$$A_k^{(m)} \leqq \tfrac{1}{2}(A_{k-1}^{(m)} + A_{k+1}^{(m)}); \tag{5}$$

which states that the differences $A_k - A_{k+1}$ $(k = 1, \cdots, n-1)$ are non-decreasing for increasing k. From this and a simple result we can deduce (4). Also Theorem 2 establishes that

$$\sum_{i=0}^{2l+1} (-1)^i \binom{m+i-1}{i} S_{m+i} \leqq p_m(1, \cdots, n) \leqq \sum_{i=0}^{2l} (-1)^i \binom{m+i-1}{i} S_{m+i}, \tag{6}$$

[2] Loc. cit., Vol. 208(1939), p 1703.

for $2l + 1 \leqq n - m$ and $2l \leqq n - m$ respectively. These inequalities throw light on formula (3) and are sharper than the following analogue of Boole's inequality for $p_m(1, \cdots, n)$, which is a special case of (4):

$$p_m(1, \cdots, n) \leqq \Sigma p_{\nu_1 \cdots \nu_m}. \tag{7}$$

The last statement will be evident in the proof.

In Theorem 5 we give an "inversion" of the formula (3), i.e we express $p_{1\cdots n}$ in terms of the $p_m(\nu_1, \cdots, \nu_k)$'s, as follows:

$$\begin{aligned}\binom{n-1}{m-1} p_{1\cdots n} &= \sum p_m(\nu_1, \cdots, \nu_m) - \sum p_m(\nu_1, \cdots, \nu_{m+1}) + \cdots \\ &\qquad + (-1)^{n-m} p_m(1, \cdots, n) \\ &= \sum_{i=0}^{n-m} (-1)^i \sum p_m(\nu_1, \cdots, \nu_{m+i}).\end{aligned} \tag{8}$$

This of course implies the following more general formula for $p_{\alpha_1 \cdots \alpha_r}$,

$$\binom{r-1}{m-1} p_{\alpha_1 \cdots \alpha_r} = \sum_{i=0}^{r-m} (-1)^i \sum p_m(\nu_1, \cdots, \nu_{m+i})$$

where $(\alpha_1, \cdots, \alpha_r)$ is a combination of the integers $(1, \cdots, n)$ and where the second summation extends to all the $\binom{r}{m+i}$-combinations of $(\alpha_1, \cdots, \alpha_r)$. Since it is known[3] that we can express other functions such as S_r, $p_{[\mu_1 \cdots \mu_r]}$ in terms of the $p_{\mu_1 \cdots \mu_r}$'s, we can also express them in terms of the $p_m(\nu_1, \cdots, \nu_k)$'s, provided $r \geqq m$.

Finally, for the case $m = 1$, we give in Theorem 6 an explicit formula for $p_{[1 \cdots r]}$ in terms of the $p_1(\nu_1, \cdots, \nu_k)$'s, as shown in (9),

$$\begin{aligned}p_{[1 \cdots r]} &= -p_1(r+1, \cdots, n) + \sum_{\nu_1} p_1(\nu_1, r+1, \cdots, n) \\ &\qquad - \sum_{\nu_1, \nu_2} p_1(\nu_1, \nu_2, r+1, \cdots, n) + \cdots \\ &\qquad + (-1)^{r-1} \sum p_1(1, \cdots, r, r+1, \cdots, n), \\ &= \sum_{i=1}^{r} (-1)^{i-1} \sum_{(\nu_1, \cdots, \nu_i)} p_1(\nu_1, \cdots, \nu_i, r+1, \cdots, n),\end{aligned} \tag{9}$$

where $(\nu_1, \cdots, \nu_i)$ runs through all the $\binom{r}{i}$-combinations from $(1, \cdots, r)$. This of course implies the following more general formula:

$$p_{[\alpha_1 \cdots \alpha_r]} = \sum_{i=1}^{r} (-1)^{i-1} \sum_{(\nu_1, \cdots, \nu_i)} p_1(\nu_1, \cdots, \nu_i, \alpha_{r+1}, \cdots, \alpha_n),$$

[3,4] Fréchet, "Condition d'existence de systemes d'événements associés à certaines probabilités," *Jour. de Math.*, (1940), p. 51–62.

where $(\alpha_1, \cdots, \alpha_r, \cdot \cdot \alpha_n)$ is a permutation of $(1, \cdots, n)$ and where $(\nu_1, \cdots, \nu_i)$ runs through all the $\binom{r}{i}$-combinations out of $(\alpha_1, \cdots, \alpha_r)$. From Theorem 6 and two lemmas we deduce a condition of existence of systems of events associated with the probabilities $p_1(\nu_1, \cdots, \nu_m)$. The author has not been able to obtain similar elegant results for the general m Probably they do not exist.

2. Generalization of Poincaré's formula; Generalization and sharpening of Boole's inequality.

THEOREM 1:

$$p_m(1, \cdots, n) = \sum p_{\nu_1 \cdots \nu_m} - \binom{m}{1} \sum p_{\nu_1 \cdots \nu_{m+1}} + \binom{m+1}{2} \sum p_{\nu_1 \cdots \nu_{m+2}} - \cdots + (-1)^{n-m} \binom{n-1}{n-m} p_{1 \cdots n}. \tag{3}$$

PROOF: We have

$$p_m(1, \cdots, n) = \sum_{b=0}^{n-m} \sum p_{[\mu_1 \cdots \mu_{m+b}]}, \tag{10}$$

where the second summation extends, for a fixed b, to all the $\binom{n}{m+b}$-combinations of $(1, \cdots, n)$. Further we have

$$p_{\nu_1 \cdots \nu_{m+c}} = \sum_{d=0}^{n-m-c} \sum p_{[\nu_1 \cdots \nu_{m+c} \cdot \nu_{m+c+d}]} \tag{11}$$

where the second summation extends, for a fixed d, to all the $\binom{n-m-c}{d}$-combinations of $(1, \cdots, n) - (\nu_1, \cdots, \nu_{m+c})$. The formulas (10) and (11) are evident by observing that the probabilities in the summations are all additive. Now we count the number of times a fixed $p_{[\mu_1 \cdots \mu_{m+b}]}$ appears in (3). By (11) this is equal to the sum

$$\binom{m+b}{m} - \binom{m}{1}\binom{m+b}{m+1} + \binom{m+1}{2}\binom{m+b}{m+2} - \cdots + (-1)^{n-m}\binom{n-1}{n-m}\binom{m+b}{m+b} = 1,$$

since this number is the coefficient of $(-1)^m x^m$ in the expansion of

$$(1-x)^{m+b}\left(1 - \frac{1}{x}\right)^{-m} = (-1)^{-m} x^m (1-x)^b.$$

Thus by (10) we have (3).

THEOREM 2: *For $2l \leqq n - m$ and $2l \leqq n - m$ respectively, we have*

$$(6)\quad \sum_{i=0}^{2l+1} (-1)^i \binom{m+i-1}{i} S_{m+i} \leqq p_m(1, \cdots, n) \leqq \sum_{i=0}^{2l} (-1)^i \binom{m+i-1}{i} S_{m+i}.$$

PROOF: By the reasoning in the previous proof, it is sufficient (in fact also necessary) to show that

$$\sum_{i=0}^{2l} \binom{m-1+i}{i}\binom{m+b}{m+i} \geqq 1, \qquad \sum_{i=0}^{2l+1} \binom{m-1+i}{i}\binom{m+b}{m+i} < 1.$$

Since

$$\binom{m-1+i}{i}\binom{m+b}{m+i} = \frac{(m+b)!}{(m-1)!\,b!}\binom{b}{i}\frac{1}{m+i}$$

is an integer, it is sufficient to show that

$$(12)\qquad \sum_{i=0}^{2l} (-1)^i \binom{b}{i}\frac{1}{m+i} > 0, \qquad \sum_{i=0}^{2l+1} (-1)^i \binom{b}{i}\frac{1}{m+i} \leqq 0.$$

Suppose $b > 0$ is even. For $i \leqq b/2 - 1$, we have $\dfrac{b-i}{i+1} > 1$ so that $\dfrac{b-i}{i+1} \geqq \dfrac{i+2}{i+1}$. Also $\dfrac{m+i}{m+i+1} \geqq \dfrac{i+1}{i+2}$ for $m \geqq 1$. Hence

$$\binom{b}{i+1}\frac{1}{m+i+1} = \frac{b-i}{i+1}\frac{m+i}{m+i+1}\binom{b}{i}\frac{1}{m+i}$$

$$\geqq \frac{i+2}{i+1}\frac{i+1}{i+2}\binom{b}{i}\frac{1}{m+i} = \binom{b}{i}\frac{1}{m+i}.$$

For $i \geqq b/2$ we have $\dfrac{b-i}{i+1} < 1$ so that $\dfrac{b-i}{i+1}\dfrac{m+i}{m+i+1} < 1$ and

$$\binom{b}{i+1}\frac{1}{m+i+1} < \binom{b}{i}\frac{1}{m+i}.$$

Thus the absolute values of the terms of the alternating series

$$\sum_{i=0}^{b} (-1)^i \binom{b}{i}\frac{1}{m+i} = \frac{b!}{(m+b)!(m-1)!}$$

are monotone increasing as long as $i \leqq \dfrac{b}{2} - 1$, reaching maximum at $i = \dfrac{b}{2}$ and then become monotone decreasing.

Therefore (12) evidently holds for $2l \leqq b/2$ and $2l + 1 \leqq b/2$ respectively. For $t \geqq \dfrac{b}{2} + 1$ we write

$$\sum_{i=0}^{t} (-1)^i \binom{b}{i}\frac{1}{m+i} = \frac{b!}{(m+b)!(m-1)!} - \sum_{i=t+1}^{b} (-1)^i \binom{b}{i}\frac{1}{m+i}$$

$$= \frac{b!}{(m+b)!(m-1)!} - \sum_{j=0}^{b-t-1} (-1)^j \binom{b}{j}\frac{1}{m+b-j}.$$

From the above and the fact that $\frac{b!}{(m+b)!(m-1)!} \leqq \frac{1}{m+b}$ we see that the righthand side is an alternating series whose terms are non-decreasing in absolute values. Hence (12) is true.

If b is odd, the case is similar.

3. Generalization of Fréchet's inequalities and related inequalities. Before proving our remaining theorems, we shall give a more detailed account of the general method which will be used. In the foregoing work we have already given two different expressions for the function $p_m(1, \cdots, n)$, namely, formulas (3) and (10), but they are not convenient for our later purposes. Formula (3) is inconvenient because it is not additive and because the $p_{\nu_1 \cdots \nu_t}$'s are related in magnitudes; while formula (10) has gone so far in the separation of the additive constituents that its application raises algebraical difficulties. Let us therefore take an intermediate course.

Let each $\binom{n}{m}$-combination $(\nu_1, \cdots, \nu_m)$ out of $(1, \cdots, n)$ be written so that $\nu_1 < \nu_2 < \cdots < \nu_m$. Then we arrange them in an ordered sequence in the following way: the combination $(\nu_1, \cdots, \nu_m)$ is to precede the combination $(\mu_1, \cdots, \mu_m)$ if, for the first $\nu_i \neq \mu_i$, we have $\nu_i > \mu_i$. After such an arrangement we symbolically denote these combinations by

$$\text{I, II, } \cdots, \left[\binom{n}{m}\right].$$

Further, all the $\binom{k}{m}$-combinations out of $(\nu_1, \cdots, \nu_k)$ where the latter is a combination out of $(1, \cdots, n)$ are arranged in the order in which they appear in the sequence just written. For example, all the $\binom{4}{2}$-combinations out of (1, 2, 3, 4) are ordered thus:

$$(12) \quad (13) \quad (14) \quad (23) \quad (24) \quad (34).$$

Let U denote a typical combination $(\mu_1, \cdots, \mu_m)$. By E_U we mean the combination of events $E_{\mu_1}, \cdots, E_{\mu_m}$ so that $p_U = p_{\mu_1 \cdots \mu_m}$. In general, let the combinations $U_1, \cdots, U_{b-1}, U_b$ be given, then $p_{U_1' \cdots U_{b-1}' U_b}$ denotes the probability of the non-occurrence of $U_1, \cdots, U_{b-1}$ and the occurrence of U_b.

Now let $\text{I, II, } \cdots, \left[\binom{k}{m} - 1\right] = Y, \left[\binom{k}{m}\right] = Z$ denote all the $\binom{k}{m}$-combinations out of $(\nu_1, \cdots, \nu_k)$ in their assigned order. We have

$$p_m(\nu_1, \cdots, \nu_k) = p_{\text{I}} + p_{\text{I}'\text{II}} + p_{\text{I}'\text{II}'\text{III}} + \cdots + p_{\text{I}' \cdots Y'Z}. \tag{13}$$

This fundamental formula is evident. Of course it is possible to identify the p's on the right-hand side with the ordinary $p_{\nu_1' \cdots \nu_j'}$'s, but we shall refrain from so doing and be content with the following example:

$$p_2(1, 2, 3, 4) = p_{12} + p_{12'3} + p_{12'3'4} + p_{1'23} + p_{1'23'4} + p_{1'2'34}.$$

THEOREM 3. *For $k = 1, \cdots, n-1$ and $1 \leqq m \leqq k$ we have*

$$\binom{n-m}{k-m} \Sigma p_m(\nu_1, \cdots, \nu_{k+1}) \leqq \binom{n-m}{k+1-m} \Sigma p_m(\nu_1, \cdots, \nu_k).$$

PROOF. Substitute (13) and a similar formula for $k+1$ into the two sides respectively. After this substitution we observe that the number of terms is the same on both sides, since

$$\binom{n-m}{k-m}\binom{n}{k+1}\binom{k+1}{m} = \binom{n-m}{k+1-m}\binom{n}{k}\binom{k}{m}.$$

Also, the number of terms with a given $U = (\mu_1, \cdots, \mu_m)$ unaccented is the same, since

$$\binom{n-m}{k-m}\binom{n-m}{k+1-m} = \binom{n-m}{k+1-m}\binom{n-m}{k-m}.$$

Let the sum of all the terms with U unaccented in the two summations be denoted by $\sigma_{k+1} = \sigma_{k+1}(\mu_1, \cdots, \mu_m)$ and $\sigma_k = \sigma_k(\mu_1, \cdots, \mu_m)$ respectively. It is sufficient to prove that

$$\binom{n-m}{k-m}\sigma_{k+1} \leqq \binom{n-m}{k+1-m}\sigma_k, \tag{14}$$

for any U. σ_k contains $\binom{n-m}{k-m}$ terms each of the form $p_{\nu_1'\cdots\nu_l'\mu_1\cdots\mu_m}$ where $0 \leqq l \leqq \mu_m - m$ and where $(\nu_1, \cdots, \nu_l, \mu_1, \cdots, \mu_m)$ is a $\binom{\mu_m}{m+l}$-combination out of $(1, \cdots, \mu_m)$. For fixed $(\mu_1, \cdots, \mu_m)$ and a fixed l but varying λ's, σ_k contains $\binom{n-\mu_m}{k-m-l}$ terms of the form $p_{\nu_1'\cdots\nu_l'\mu_1\cdots\mu_m}$, with exactly l accented subscripts. Let the sum of all such terms be denoted by $\sigma_k^{(l)}$. Evidently $\sigma_k^{(l)}$ has $\binom{\mu_m - m}{l}$ terms. As a check we have

$$\binom{n-\mu_m}{k-m}\binom{\mu_m-m}{0} + \binom{n-\mu_m}{k-m-1}\binom{\mu_m-m}{1} + \cdots + \binom{n-\mu_m}{k-\mu_m}\binom{\mu_m-m}{\mu_m-m} = \binom{n-m}{k-m},$$

which is the total number of terms in σ_k.

We decompose these p's partially, as follows:

$$p_{\nu_1'\cdots\nu_l'\mu_1\cdots\mu_m} = \sum_{b=0}^{\mu_m-m-l} \sum_{\mu_{m+1}\cdots\mu_{m+b}} p_{\nu_1'\cdots\nu_{l+c}'\mu_1\cdots\mu_{m+b}},$$

where $(\nu_1, \cdots, \nu_{l+c}, \mu_1, \cdots, \mu_{m+b})$ is a permutation of $(1, \cdots, \mu_m)$ and where the second summation extends, for a fixed b, to all the $\binom{\mu_m - m - l}{b}$-combinations out of $(1, \cdots, \mu_m) - (\nu_1, \cdots, \nu_l, \mu_1, \cdots \mu_m)$.

Now consider a given

$$p_{\rho i \cdots \rho_t \lambda_1 \cdots \lambda_s \mu_1 \cdots \mu_m}$$

where $0 \leqq t \leqq \mu_m - m$ and $(\rho_1 \cdots \rho_t \lambda_1 \cdots \lambda_s \mu_1 \cdots \mu_m)$ is a permutation of $(1, \cdots, \mu_m)$. It appears $\binom{t}{l}$ times in $\sigma_k^{(l)}$. Hence it appears

$$\binom{n-\mu_m}{k-m}\binom{t}{0} + \binom{n-\mu_m}{k-m-1}\binom{t}{1} + \cdots + \binom{n-\mu_m}{k-m-t}\binom{t}{t} = \binom{n-\mu_m+t}{k-m}$$

times in σ_k.

Therefore to prove (14) it is sufficient to prove that

$$\binom{n-m}{k-m}\binom{n-\mu_m+t}{k+1-m} \leqq \binom{n-m}{k+1-m}\binom{n-\mu_m+t}{k-m}.$$

By an easy reduction we have

$$(n - \mu_m + t - k + m) \leqq n - k$$

or

$$-\mu_m + t + m \leqq 0;$$

since $t \leqq \mu_m - m$ this is obvious.

THEOREM 4: *For $2 \leqq k \leqq n-1$ and $1 \leqq m \leqq k$ we have*

$$(5)_a \qquad \frac{\Sigma p_m(\nu_1, \cdots, \nu_k)}{\binom{n-m}{k-m}} \leqq \frac{1}{2}\frac{\Sigma p_m(\nu_1, \cdots, \nu_{k-1})}{\binom{n-m}{k-1-m}} + \frac{1}{2}\frac{\Sigma p_m(\nu_1, \cdots, \nu_{k+1})}{\binom{n-m}{k+1-m}}.$$

PROOF: By the reasoning in the previous proof, it is sufficient to show that

$$2\binom{n-m}{k-1-m}\binom{n-m}{k+1-m}\binom{n-\mu_m+t}{k-m} \leqq \binom{n-m}{k-m}\binom{n-m}{k+1-m}\binom{n-\mu_m+t}{k-1-m} + \binom{n-m}{k-m}\binom{n-m}{k-1-m}\binom{n-\mu_m+t}{k+1-m},$$

for $0 \leqq t \leqq \mu_m - m$. By an easy reduction this is equivalent to

$$2(n-k)(n-\mu_m+t-k+m+1) \leqq (n-k+1)(n-k) + (n-\mu_m+t-k+m+1)(n-\mu_m+t-k+m)$$

or

$$(n-\mu_m+t-k+m+1)(\mu_m-t-m) \leqq (n-k)(\mu_m-t-m).$$

For $t = \mu_m - m$ we have equality, otherwise we have

$$-\mu_m + t + m + 1 \leqq 0.$$

We can deduce Theorem 3 from Theorem 4 and the following result (a case of generalized Gumbel inequalities):

$$\binom{n-1}{n-2} p_m(1, \cdots, n) \leqq \Sigma p_m(\nu_1, \cdots, \nu_{n-1}). \tag{15}$$

PROOF OF (15). Substitute from (13). Consider the p's with U unaccented. The number of such terms is the same on both sides. But on the left-hand side they are all the same $p_{12\cdots(v-1)v}$, while those on the right-hand side, being of the form $p_{v\lambda}$ $_{v\lambda v}$ where $0 \leqq \lambda \leqq U - 1$ and $(U_1, \cdots, U_\lambda)$ is a combination out of $(1, \cdots, U-1)$, are greater than or equal to it. Hence the result.

4. The $p_{\alpha_1\cdots\alpha_s}$'s in terms of the $p_m(\nu_1, \cdots, \nu_k)$'s and the $p_{(\alpha_1\cdots\alpha_s)}$'s in terms of the $p_1(\nu_1, \cdots, \nu_k)$'s.

THEOREM 5: *For $1 \leqq m \leqq n$ we have*

$$\begin{aligned}\binom{n-1}{m-1} p_{1\cdots n} &= \sum p_m(\nu_1, \cdots, \nu_m) - \sum p_m(\nu_1, \cdots, \nu_{m+1}) + \cdots \\ &\qquad + (-1)^{n-m} p_m(1, \cdots, n) \\ &= \sum_{i=0}^{n-m} (-1)^i \sum_{\nu_1, \cdots \nu_{m+i}} p_m(\nu_1, \cdots, \nu_{m+i}).\end{aligned} \tag{8}$$

PROOF: As in the proof of Theorem 3, consider $\sigma_k(\mu_1, \cdots, \mu_m)$. Here $m \leqq k \leqq n$. Since a given

$$p_{\rho_1\cdots\rho_t\lambda_1\cdots\lambda_s\mu_1\cdots\mu_m}, \tag{16}$$

appears $\binom{n-\mu_m+t}{k-m}$ times in σ_k, it appears

$$\begin{aligned}\sum_{k=m}^{n} (-1)^{k-m}\binom{n-\mu_m+t}{k-m} &= \sum_{j=0}^{n-m} (-1)^j \binom{n-\mu_m+t}{j} \\ &= \sum_{j=0}^{n-\mu_m+t} (-1)^j \binom{n-\mu_m+t}{j} = \begin{cases} 0, & \text{if } n-\mu_m+t \geqq 1, \\ 1, & \text{if } n-\mu_m+t = 0. \end{cases}\end{aligned}$$

times on the right hand side of (8). Hence for fixed $(\mu_1, \cdots, \mu_m)$, the only p's of the form (16) which actually appears are those with $t = \mu_m - n$. But $\mu_m \leqq n$, thus $t = 0$, $\mu_m = n$, and $(\lambda_1, \cdots \lambda_s, \mu_1, \cdots, \mu_m)$ is a permutation of $(1, \cdots, n)$. The term in question is therefore $p_{1\cdots n}$. Since the number of $\binom{n}{m}$-combinations of $(1, \cdots, n)$ with $\mu_m = n$ is $\binom{n-1}{m-1}$, we have the theorem.

THEOREM 6: *For $1 \leqq r \leqq n-1$, we have*

$$\begin{aligned}p_{[1\cdots r]} &= -p_1(r+1, \cdots, n) + \sum_{\nu_1} p_1(\nu_1, r+1, \cdots, n) \\ &\quad - \sum_{\nu_1,\nu_2} p(\nu_1, \nu_2, r+1, \cdots, n) + \cdots + (-1)^{r-1} \sum p_1(1, \cdots, n) \\ &= \sum_{i=1}^{r} (-1)^{i-1} \sum_{\nu_1, \cdots, \nu_i} p_1(\nu_1, \cdots, \nu_i, r+1, \cdots, n),\end{aligned} \tag{9}$$

where $(\nu_1, \cdots, \nu_i)$ *runs through all the* $\binom{r}{i}$*-combinations out of* $(1, \cdots, r)$

PROOF: We rewrite (14) for the special case $m = 1$,

$$p_1(\mu_1, \cdots, \mu_k) = p_{\mu_1} + p_{\mu_1'\mu_2} + \cdots + p_{\mu_1' \cdots \mu_{k-1}'\mu_k}, \tag{17}$$

where $\mu_1 < \mu_2 < \cdots < \mu_k$. Substitute into the right hand side of (9). After the substitution let the sum of all those p's with μ unaccented be denoted by σ_μ. The terms in σ_μ are of the form $p_{\mu_1' \cdots \mu_{s-1}'\mu}$ where $1 \leqq s \leqq \mu$ and $(\mu_1, \cdots, \mu_{s-1})$ is a combination out of $(1, \cdots, \mu - 1)$.

First consider a fixed $\mu \leqq r$. For a fixed $p_{\mu_1' \cdots \mu_{s-1}'\mu}$ we count the number of times it appears in σ_μ, that is, on the right hand side of (9) This is evidently equal to

$$\sum_{j=s}^{r} (-1)^j \binom{r-\mu}{j-s} = \sum_{j=s}^{r-\mu+s} (-1)^j \binom{r-\mu}{j-s} = \begin{cases} 0, & \text{if } r - \mu \geqq 1, \\ 1, & \text{if } r - \mu = 0. \end{cases}$$

Thus the only terms that actually appear are those with $\mu = r$; and each of such terms $p_{\mu_1' \cdots \mu_{s-1}'r}$ appears exactly once with the sign $(-1)^s$. Hence their total contribution is

$$p_r = \sum_{\nu_1} p_{\nu_1'r} + \sum_{\nu_1, \nu_2} p_{\nu_1'\nu_2'r} - \cdots + (-1)^{r-1} p_{1' \cdots (r-1)'r} = p_{1 \cdots r}, \tag{18}$$

by an easy modification of Poincaré's formula.

Next consider a fixed $\mu \geqq r + 1$. Every term with μ unaccented in σ_μ is of the form (with the usual convention for $\mu = r + 1$) $p_{\mu_1' \cdots \mu_s'(r+1)' \cdots (\mu-1)'\mu}$, where $(\mu_1, \cdots, \mu_s)$ is a combination out of $(1, \cdots, r)$; and it appears exactly once with the sign $(-1)^s$. Their total contribution is therefore

$$-p_{(r+1)' \cdots (\mu-1)'\mu} + \sum_{\nu_1} p_{\nu_1'(r+1)' \cdots (\mu-1)'\mu} - \sum_{\nu_1, \nu_2} p_{\nu_1'\nu_2'(r+1)' \cdots (\mu-1)'\mu} + \cdots$$
$$+ (-1)^{r-1} p_{1' \cdots (\mu+1)'\mu} = -p_{1 \cdots r(r+1)' \cdots (\mu-1)'\mu},$$

by another application of Poincaré's formula. Summing up for $\mu = r + 1, \cdots, n$, we obtain

$$-(p_{1 \cdots r(r+1)} + p_{1 \cdots r(r+1)'(r+2)} + \cdots + p_{1 \cdots r(r+1)' \cdots (n-1)'n}). \tag{19}$$

Adding (18) and (19), we obtain as the sum of the right-hand side of (9)

$$p_{1 \cdots r} - (p_{1 \cdots r(r+1)} + p_{1 \cdots r(r+1)'(r+2)} + \cdots + p_{1 \cdots r(r+1)' \cdots (n-1)'n})$$
$$= p_{1 \cdots (r+1)'(r+2)' \cdots n'} = p_{[1 \cdots r]}$$

by an easy modification of (17).

5. A condition for existence of systems of events associated with the probabilities $p_1(\nu_1, \cdots, \nu_k)$.

LEMMA 1: *Let any* $2^n - 1$ *quantities* $q(\alpha_1, \cdots, \alpha_k)$ *be given, where* $k =$

$1, \cdots, n$, *and for a fixed* k, $(\alpha_1, \cdots, \alpha_k)$ *runs through all the* $\binom{n}{k}$*-combinations out of* $(1, \cdots, n)$. *Let the quantities* $Q(\alpha_1, \cdots, \alpha_k)$ *be formed as follows*:

$$Q(0) = 1 - q(1, \cdots, n),$$

$$Q(\alpha_1, \cdots, \alpha_k) = -q(\alpha_{k+1}, \cdots, \alpha_n) + \sum_{\nu_1} q(\nu_1, \alpha_{k+1}, \cdots, \alpha_n) - \sum_{\nu_1, \nu_2} q(\nu_1, \nu_2, \alpha_{k+1}, \cdots, \alpha_n) + \cdots + (-1)^{k-1} q(1, \cdots, n),$$

where $(\nu_1, \cdots, \nu_i)$ runs through all the $\binom{k}{i}$-combinations out of $(1, \cdots, n) - (\alpha_{k+1}, \cdots, \alpha_n)$. *Then the sum of all these* Q*'s is equal to* 1.

PROOF: Add all these Q's and count the number of times a fixed $q(\mu_1, \cdots, \mu_k)$ appears in the sum. For $1 \leqq k \leqq n$ this number is equal to

$$-1 + \binom{k}{1} - \binom{k}{2} + \cdots + (-1)^{k-1}\binom{k}{k} = 0.$$

Hence we have the lemma.

LEMMA 2: (Fréchet) *Given* 2^n *quantities* $Q_{[\alpha_1 \cdots \alpha_r]}$ *where* $(\alpha_1, \cdots, \alpha_r)$ *runs through all combinations out of* $(1, \cdots, n)$ *including the empty one. The necessary and sufficient condition that there exist systems of events* $E_1, \cdots, E_n$ *for which*

$$p_{[\alpha_1 \cdots \alpha_r]} = Q_{[\alpha_1 \cdots \alpha_r]}$$

(where $p_{[0]}$ denotes the probability for the non-occurrence of $E_1, \cdots, E_n$) *is that each* $Q \geqq 0$ *and that their sum is equal to* 1.

PROOF: Since the probabilities $p_{[\alpha_1 \cdots \alpha_r]}$ are independent, i.e., unrelated in magnitudes except that their sum is equal to 1, the lemma is evident.

THEOREM 7: *Given* $2^n - 1$ *quantities* $q(\alpha_1, \cdots, \alpha_k)$ *as in* Lemma 1, *the necessary and sufficient condition that there exist systems of events* $E_1, \cdots, E_n$ *for which*

$$p_1(\alpha_1, \cdots, \alpha_k) = q(\alpha_1, \cdots, \alpha_k)$$

is that for any combination $(\alpha_{r+1}, \cdots, \alpha_n)$, $1 \leqq r \leqq n-1$, *out of* $(1, \cdots, n)$ *we have*

$$-q(\alpha_{r+1}, \cdots, \alpha_n) + \sum_{\nu_1} q(\alpha_{\nu_1}, \alpha_{r+1}, \cdots, \alpha_n) - \sum_{\nu_1, \nu_2} q(\alpha_{\nu_1}, \alpha_{\nu_2}, \alpha_{r+1}, \cdots, \alpha_n) + \cdots + (-1)^{r-1} q(1, \cdots, n) \geqq 0,$$

and thus

$$1 - q(1, \cdots, n) \geqq 0.$$

PROOF: The condition is necessary by Theorem 6. It is sufficient by Lemma 1, 2 and an obvious formula expressing $p_1(\alpha_1, \cdots, \alpha_r)$ in terms of the $p_{[\nu_1 \cdots \nu_i]}$'s.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items

A NOTE ON SHEPPARD'S CORRECTIONS

By Cecil C. Craig

University of Michigan

As far as the author is aware, H C. Carver was the first to point out that while the formulae ordinarily given for Sheppard's corrections for central moments are valid for moments computed about the population mean, there are still systematic errors present when they are applied to central moments calculated from any particular grouped frequency distribution [1]. This is due, of course to the fact that the mean of a grouped frequency distribution is in general different from that of the distribution before grouping For a fixed class interval k, Sheppard's corrections give the average value of a moment about a fixed point of a given order for all the groupings of this class width possible and will fail to do so if the moment in question is calculated for each position of the class limits about a point which varies as the class limits shift. Thus Carver [1] pointed that the commonly used formula (for a continuous variate),

$$\mu_2 = \nu_2 - \frac{k^2}{12}, \tag{1}$$

should, if ν_2 is calculated about the mean of the grouped distribution as it is in practice, be replaced by

$$\mu_2 = \nu_2 - \frac{k^2}{12} + \sigma_M^2 \tag{2}$$

in which σ_M^2 is the variance of the means of grouped distributions over all positions of the class limits with the fixed class width k

Recently J. A. Pierce [2] gave a method for deriving the required formulae of the type of (2) and gave actual formulae for both moments and semiinvariants through the sixth order. It is the purpose of this note to point out that the use of moment generating functions provides a more elegant and concise way of arriving at formulae equivalent to Pierce's though in a somewhat different form. This method can be immediately extended to distributions of two or more variates.

In a previous paper [3] on Sheppard's corrections for a discrete variate, the author made use of the following argument· It is assumed that for a fixed class width k, any point in the scale on which the variate x is plotted is as likely to be

chosen as a class limit as any other; choosing a system of class limits for grouping the data is then equivalent to placing at random on the x-axis a scale with division points at intervals of k. Once the system of class limits is chosen any value of x before grouping bears to the class mark, x_i, of the class in which it falls the relation,

$$x_i = x + \epsilon, \tag{3}$$

in which x and ϵ are independent variates. The frequency law governing x, is, of course, that of the population from which it is drawn while ϵ is distributed in a rectangular distribution with the range $\left(-\frac{k}{2}, \frac{k}{2}\right)$ for a continuous variate and $\left(-\frac{m-1}{2m}k, \frac{m-1}{2m}k\right)$ if m consecutive values of a discrete variate are grouped in each class interval. In either case

$$M_{x_i}(\vartheta) = M_x(\vartheta)M_\epsilon(\vartheta) \tag{4}$$

in which $M_{x_i}(\vartheta)$ is the moment generating function of the variate x_i, etc. The expansion of both sides of (4) in powers of ϑ gives the relations between the average values of moments of the grouped distribution over all positions of the scale and the moments of the ungrouped distribution from which Sheppard's corrections are obtained by solving for the moments of the ungrouped distribution. The relations are valid for any fixed point about which the moments are computed; if this fixed point be taken as the mean of the ungrouped distribution the ordinary Sheppard's corrections for central moments result.

But it is quite easy to modify (4) to give the necessary relations in case the moments of each grouped distribution are computed about the mean of that distribution. We have only to write

$$x_i = x_i - \bar{x} + \bar{x} \tag{5}$$

in which $\bar{x}$ is the mean of the grouped distribution for which x_i is one of the class marks. Then

$$\begin{aligned} M_{x_i}(\vartheta) &= M_{x_i-\bar{x},\bar{x}}(\vartheta, \omega)\,|_{\omega=\vartheta} \\ &= M_x(\vartheta)M_\epsilon(\vartheta) \end{aligned} \tag{6}$$

If we write,

$$\lambda_{rs;k_i-\bar{x},\bar{x}} = \bar{\lambda}_{rs},$$

in which $\bar{\lambda}_{rs}$ is the product seminvariant of order rs of moments about the means of the grouped distributions and of such means, the expansion of the logarithm of the second member of (6) gives

$$1 + (\bar{\lambda}_{10} + \bar{\lambda}_{01})\vartheta + (\bar{\lambda}_{20} + 2\bar{\lambda}_{11} + \bar{\lambda}_{02})\frac{\vartheta^2}{2} + (\bar{\lambda}_{10} + \bar{\lambda}_{01})^{(3)}\frac{\vartheta^3}{3!} + \cdots, \tag{7}$$

in which

$$(\bar{\lambda}_{10} + \bar{\lambda}_{01})^{(r)} = \bar{\lambda}_{r0} + r\bar{\lambda}_{r-1,1} + \cdots + \binom{r}{k}\bar{\lambda}_{r-k,k} + \cdots + \bar{\lambda}_{0r}\,.$$

The expression of the logarithm of the right member is[3].

$$(8) \qquad \lambda_1\vartheta + \lambda_2\frac{\vartheta^2}{2!} + \lambda_3\frac{\vartheta^3}{3!} + \cdots + \sum_{s=1}^{\infty}(-1)^{s+1}\frac{B_s k^{2s}}{2s}\left(1 - \frac{1}{m^{2s}}\right)\frac{\vartheta^{2s}}{(2s)!},$$

for a discrete variate (the result for a continuous variable is obtained merely by letting $m \to \infty$) in which λ_r is the rth seminvariant of the ungrouped distribution and B_s is the sth Bernoulli number.

We may without loss of generality take the origin for x at the mean of the ungrouped distribution so that $\lambda_1 = 0$. Further it is easy to see that

$$\bar{\lambda}_{1r} = 0, \qquad r = 0, 1, 2, 3, \cdots .$$

Consider

$$E[(x_i - \bar{x})\bar{x}^r] = \bar{\nu}_{1r}$$

For a fixed $\bar{x}$, i.e , for a given grouping, this becomes

$$\bar{x}^r E(x_i - \bar{x}) = 0$$

Then since $\bar{\nu}_{1r}$ is the average of this over all groupings with a given class interval, $\bar{\nu}_{1r} = 0$, and from the expression for $\bar{\lambda}_{1r}$ in terms of the moments $\bar{\nu}_{i1}$ it is obvious that also $\bar{\lambda}_{1r} = 0$.

Then we must also have $\bar{\lambda}_{01} = 0$ as is otherwise obvious and (7) can be rewritten

$$(9) \qquad 1 + (\bar{\lambda}_{20} + \bar{\lambda}_{02})\frac{\vartheta^2}{2} + (\bar{\lambda}_{30} + 3\bar{\lambda}_{21} + \bar{\lambda}_{03})\frac{\vartheta^3}{3!} + \cdots .$$

Now from (8) and (9) by equating coefficients of like powers of ϑ, we get the set of formulae:

$$(10) \qquad \begin{aligned} \lambda_1 &= 0 \\ \lambda_2 &= \bar{\lambda}_{20} + \bar{\lambda}_{02} - \left(1 - \frac{1}{m^2}\right)\frac{k^2}{12} \\ \lambda_3 &= \bar{\lambda}_{30} + 3\bar{\lambda}_{21} + \bar{\lambda}_{03} \\ \lambda_4 &= \bar{\lambda}_{40} + 4\bar{\lambda}_{31} + 6\bar{\lambda}_{22} + \bar{\lambda}_{04} + \left(1 - \frac{1}{m_4}\right)\frac{k^4}{120} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}$$

These formulae, however, do not give the sought Sheppard's corrections for seminvariants calculated from grouped distributions of a discrete variate. See below.

Referring to formula (10), p. 58 of the author's paper cited [3], it is easily seen by comparison that the required moment formulae are obtained from the general formula

$$(11) \qquad \mu_n = \sum_{s=0}^{[n/2]}\binom{n}{2s}\alpha_{2s}(\bar{\nu}_{10} + \bar{\nu}_{01})^{(n-2s)},$$

in which α_{2s} is given by formula (9) of this former paper. For $n = 1, 2, 3, 4$ we write down immediately

$$\mu_1 = 0 \qquad (\bar{\nu}_{10} = \bar{\nu}_{01} = 0)$$

$$\mu_2 = \bar{\nu}_{20} + \bar{\nu}_{02} - \left(1 - \frac{1}{m^2}\right)\frac{k^2}{12}$$

$$\mu_3 = \nu_{30} + 3\bar{\nu}_{21} + \bar{\nu}_{03} \tag{12}$$

$$\mu_4 = \nu_{40} + 4\bar{\nu}_{31} + 6\bar{\nu}_{22} + \bar{\nu}_{04} - \left(1 - \frac{1}{m^2}\right)(\bar{\nu}_{20} + \bar{\nu}_{02})\frac{k^2}{2} + \left(1 - \frac{1}{m^2}\right)\left(7 - \frac{3}{m^2}\right)\frac{k^4}{240}.$$

In these formulae, $\bar{\nu}_{r0}$ is, of course, the average value of rth central moments about the means of grouped distributions. From the definition $\bar{\nu}_{rs} (s \neq 0)$ is the average value of the product of the rth central moment of a grouped distribution by the sth power of the mean of the same grouped distribution. Also, it must be noted that in the formulae (10) the $\bar{\lambda}_{rs}$'s there are to be calculated by the usual formulae from the moments, $\bar{\nu}_{ij}$, , and are not themselves the average values of like seminvariants calculated from the separate grouped distributions. Thus though the formulae (12) give the sought Sheppard's corrections for moments, the formulae (10) do not do the like for seminvariants in general. However, since in each grouped distribution,

$$\lambda_2 = \nu_2$$

and

$$\lambda_3 = \nu_3$$

we have, taking the expectation or average value over the grouped distributions,

$$E(\lambda_2) = E(\nu_2) = \bar{\nu}_{20} = \bar{\lambda}_{20}$$

and

$$E(\lambda_3) = E(\nu_3) = \bar{\nu}_{30} = \bar{\lambda}_{30},$$

and the first two formulae of (10) do give the Sheppard's corrections for λ_2 and λ_3 calculated from grouped distributions of a discrete variate.

But the case for λ_4 is different. In each grouped distribution,

$$\lambda_4 = \nu_4 - 3\nu_2^2,$$

and if we define l_r by

$$E(\lambda_r) = l_r,$$

we have

$$l_4 = \bar{\nu}_{40} - 3E(\nu_2^2)$$
$$= \bar{\nu}_{40} - 3(\bar{\nu}_{20}^2 + \nu_{2;\nu_2}) = \bar{\lambda}_{40} - 3\nu_{2\,\nu_2},$$

if $\nu_{2;\nu_2}$ is the variance of ν_2 in the grouped distributions.

In a similar way one can obtain such formulae for seminvariants as may be required. Through the sixth, the formulae for the Sheppard's corrections for the seminvariants calculated from a grouped distribution of a discrete variate are:

$$\begin{aligned}
\lambda_2 &= l_2 + \bar{\lambda}_{02} - \left(1 - \frac{1}{m^2}\right)\frac{k^2}{12} \\
\lambda_3 &= l_3 + 3\bar{\lambda}_{21} + \bar{\lambda}_{03} \\
\lambda_4 &= l_4 + 3\nu_{2,\nu_2} + 4\bar{\lambda}_{31} + 6\bar{\lambda}_{22} + \bar{\lambda}_{04} + \left(1 - \frac{1}{m^4}\right)\frac{k_4}{120} \\
\lambda_5 &= l_5 + 10\nu_{11,\nu_2,\nu_3} + 5\bar{\lambda}_{41} + 10\bar{\lambda}_{32} + 10\bar{\lambda}_{23} + \bar{\lambda}_{05} \\
\lambda_6 &= l_6 + 15\nu_{11\,\nu_2,\nu_4} + 10\nu_{2\,\nu_2} - 30\nu_{3,\nu_2} - 90\nu_{2\,\nu_2}\bar{\nu}_{20} \\
&\qquad + 6\bar{\lambda}_{51} + 15\bar{\lambda}_{42} + 20\bar{\lambda}_{33} + 15\bar{\lambda}_{24} + \bar{\lambda}_{06} - \left(1 - \frac{1}{m^6}\right)\frac{k^6}{252}.
\end{aligned} \tag{13}$$

In these formulae, $\nu_{ij\,\nu_r,\nu_s}$ is the ijth central product moment of ν_r and ν_s in the grouped distributions.

To illustrate these formulae numerically and to facilitate comparison with Pierce's results, we will use the example he chose. His ungrouped distribution was:

v	f	v	f	v	f
1	2	4	30	7	1
2	8	5	4	8	1
3	10	6	3	9	1

From this the following three grouped distributions with $k = 3$ can be formed:

(1)		(2)		(3)	
class	f	class	f	class	f
1– 3	20	0– 2	10	−1 [−1]	2
4–	37	3–	44	2–	48
7–	3	6–	5	5–	8
10–12	0	9–11	1	8–10	2

With origin at $\nu = 4$, we have the following table of moment characteristics of these four distributions:

Distribution	ν_1'	$\nu_2 = \bar{\lambda}_2$	$\nu_3 = \bar{\lambda}_3$	ν^4	$\bar{\lambda}^4$	$\delta\nu_1' = \nu_1' - \left(-\frac{10}{60}\right)$
(1)	$\frac{9}{60}$	$\frac{9819}{60^2}$	$\frac{17442}{60^3}$	$\frac{238849317}{60^4}$	$\frac{50388966}{60^4}$	$\frac{19}{60}$
(2)	$-\frac{9}{60}$	$\frac{10179}{60^2}$	$\frac{567162}{60^3}$	$\frac{557840277}{60^4}$	$\frac{247004154}{60^4}$	$\frac{1}{60}$
(3)	$-\frac{30}{60}$	$\frac{8820}{60^2}$	$\frac{1317600}{60^3}$	$\frac{528282000}{60^4}$	$\frac{294904800}{60^4}$	$-\frac{20}{60}$
Average	$-\frac{10}{60}$	$\frac{9606}{60^2}$	$\frac{622440}{60^3}$	$\frac{441657198}{60^4}$	$\frac{103839996}{60^4}$	
	μ_1'	$\mu_2 = \lambda_2$	$\mu_3 = \lambda_3$	μ_4	λ_4	
Original Distribution	$-\frac{10}{60}$	$\frac{7460}{60^2}$	$\frac{642400}{60^3}$	$\frac{305034000}{60^4}$	$\frac{138079200}{60^4}$	

From the table,

$$\bar{\nu}_{20} = \bar{\lambda}_{20} = \frac{9606}{60^2}$$

$$\bar{\nu}_{30} = \bar{\lambda}_{30} = \frac{622440}{60^3}$$

$$\bar{\nu}_{40} = \bar{\lambda}_{40} + 3\bar{\lambda}_{20}^2 = \frac{441657198}{60^4}.$$

We further compute:

$$\bar{\nu}_{02} = \frac{\Sigma(\delta\nu_1')^2}{3} = \frac{254}{60^2} = \bar{\lambda}_{02} \qquad \bar{\nu}_{21} = \frac{\Sigma(\nu_2\delta\nu_1')}{3} = \frac{6780}{60^3} = \bar{\lambda}_{21}$$

$$\bar{\nu}_{03} = \frac{-380}{60^3} = \bar{\lambda}_{03} \qquad \bar{\nu}_{30} = -\frac{8705412}{60^4} = \bar{\lambda}_{31}$$

$$\bar{\nu}_{04} = \frac{96774}{60^4} \qquad \bar{\nu}_{22} = \frac{2360946}{60^4}$$

$$\bar{\lambda}_{22} = \bar{\nu}_{22} - \bar{\nu}_{20}\bar{\nu}_{02} = \frac{-72978}{60^4}$$

$$\bar{\lambda}_{04} = \bar{\nu}_{04} - 3\bar{\nu}_{02}^2 = \frac{-96774}{60^4}$$

$$\nu_{2,\nu_2} = \frac{\Sigma\nu_2^2}{3} - \bar{\nu}_{20}^2 = \frac{330948}{60^4}$$

$$l_4 = \tilde{\lambda}_{40} - 3\nu_{2\,\nu_2} = \bar{\nu}_{40} - 3\bar{\nu}_{20}^2 - 3\nu_{2,\nu_1} = \frac{163839996}{60^4}$$

$$\left(1 - \frac{1}{m^2}\right)\frac{k^2}{12} = \frac{2}{3}$$

$$\left(1 - \frac{1}{m^2}\right)\left(7 - \frac{3}{m^2}\right)\frac{k^4}{240} = 2.$$

With these values one may check the formulae (12) and (13) as far as weight four. For example:

$$\mu_2 = \frac{9606}{60^2} + \frac{254}{60^2} - \frac{2}{3} = \frac{7460}{60^2}$$

$$\lambda_4 = \frac{1}{60^4}(163839996 + 991494 - 34821648 - 437868 - 96774 + 8640000)$$

$$= \frac{138079200}{60^4}.$$

It may appear at first glance that since

$$\bar{\nu}_{rs} = E[\nu_r(\delta\nu_1)^s]$$

and could be expressed by means of the notation, $\nu_{1^s\cdot\nu_r,\nu_1}{}'$, the notation in (12) and (13) could be made more uniform. It could be but at the expense of greater complexity in these two sets of results. Moreover, it is convenient that $\tilde{\lambda}_{rs}$ is expressible in terms of $\bar{\nu}_{kl}$'s in precisely the same way that product seminvariants are ordinarily expressible in terms of product moments.

Pierce's results differ from the above not only in their mode of derivation but also in the fact that they express $\bar{\nu}_{r0}$'s and l_r's in terms of the characteristics of the ungrouped distribution and moments and seminvariants of moments in the grouped distributions. Thus as they stand they are not formulae for Sheppard's corrections

Finally it must be remarked that in comparison with the usual formulae for Sheppard's corrections, the formulae (10) and (13) introduce quantities the magnitudes of which are not known in general except that ordinarily they are quite small. It is hoped that results on this point will be forthcoming soon.

REFERENCES

[1] H. C Carver, "The fundamental nature and proof of Sheppard's adjustments," *Annals of Math. Stat*, Vol 7 (1936), pp 154–163.

[2] J A. Pierce, "A study of a universe of n finite populations with application to moment-function adjustments for grouped data," *Annals of Math. Stat.*, Vol 11 (1940), pp. 311–334.

[3] C C Craig, "Sheppard's corrections for a discrete variable," *Annals of Math Stat.*, Vol. 7 (1936), pp. 55–61.

ON THE ANALYSIS OF VARIANCE IN CASE OF MULTIPLE CLASSIFICATIONS WITH UNEQUAL CLASS FREQUENCIES

By Abraham Wald[1]

Columbia University

In a previous paper[2] the author considered the case of a single criterion of classification with unequal class frequencies and derived confidence limits for σ'^2/σ^2 where σ'^2 denotes the variance associated with the classification, and σ^2 denotes the residual variance. The scope of the present paper is to extend those results to the case of multiple classifications with unequal class frequencies.

For the sake of simplicity of notations we will derive the required confidence limits in the case of a two-way classification, the extension to multiple classifications being obvious.

Consider a two-way classification with p rows and q columns. Let y be the observed variable, and let n_{ij} be the number of observations in the ith row and jth column. Denote by $y_{ij}^{(k)}$ the kth observation on y in the ith row and jth column ($k = 1, \cdots, n_{ij}$). Let the total number of observations be N. We order the N observations and let y_α be the αth observation on y in that order. Consider the variables:

$$t, t_1, \cdots, t_p, v_1, \cdots, v_q,$$

and denote by t_α the αth observation on t, by $t_{i\alpha}$ the αth observation on t_i and by $v_{j\alpha}$ the αth observation on v_j. The values of t_α, $t_{i\alpha}$ and $v_{j\alpha}$ are defined as follows:

$$t_\alpha = 1 \ (\alpha = 1, \cdots, N),$$

$$t_{i\alpha} = 1 \text{ if } y_\alpha \text{ lies in the } i\text{th row},$$

$$t_{i\alpha} = 0 \text{ if } y_\alpha \text{ does not lie in the } i\text{th row},$$

$$v_{j\alpha} = 1 \text{ if } y_\alpha \text{ lies in the } j\text{th column},$$

$$v_{j\alpha} = 0 \text{ if } y_\alpha \text{ does not lie in the } j\text{th column}.$$

We make the assumptions

$$y_{ij}^{(k)} = x_{ij}^{(k)} + \epsilon_i + \eta_j,$$

where the variates $x_{ij}^{(k)}$, ϵ_i, η_j ($i = 1, \cdots, p$; $j = 1, \cdots, q$; $k = 1, \cdots, n_{ij}$) are independently and normally distributed, the variance of $x_{ij}^{(k)}$ is σ^2, the variance of ϵ_i is σ'^2, the variance of η_j is σ''^2, and the mean values of ϵ_i and η_j are zero.

---

[1] Research under a grant-in-aid from the Carnegie Corporation of New York

[2] "A note on the analysis of variance with unequal class frequencies," *Annals of Math. Stat.*, Vol. 11 (1940).

Let the sample regression of y on $t, t_1, \cdots, t_{p-1}, v_1, \cdots, v_{q-1}$ be

$$Y = at + b_1t_1 + \cdots + b_{p-1}t_{p-1} + d_1v_1 + \cdots + d_{q-1}v_{q-1}.$$

We want to derive confidence limits for

$$\sigma'^2/\sigma^2 = \lambda^2.$$

Let us introduce the notations:

$$\begin{aligned}
\sum_\alpha t_\alpha t_{i\alpha} &= a_{0i} && (i = 1, \cdots, p-1),\\
\sum t_\alpha v_{j\alpha} &= a_{0p-1+j} && (j = 1, \cdots, q-1),\\
\sum t_{i\alpha} t_{j\alpha} &= a_{ij} && (i, j = 1, \cdots, p-1),\\
\sum t_{i\alpha} v_{j\alpha} &= a_{ip-1+j} && (i = 1, \cdots, p-1; j = 1, \cdots, q-1),\\
\sum v_{i\alpha} v_{j\alpha} &= a_{p-1+i\ p-1+j} && (i, j = 1, \cdots, q-1),\\
\| c_{ij} \| &= \| a_{ij} \|^{-1} && (i, j = 0, 1, \cdots, p+q-2).
\end{aligned}$$

Let the regression of $x_i^{(k)}$ on $t, t_1, \cdots, t_{p-1}, v_1, \cdots, v_{q-1}$ be

$$X = a^*t + b_1^*t_1 + \cdots + b_{p-1}^*t_{p-1} + d_1^*v_1 + \cdots + d_{q-1}^*v_{q-1}.$$

The regression of $\epsilon_i + \eta_j$ on the same independent variables is evidently equal to

$$\begin{aligned}
&\epsilon_1t_1 + \cdots + \epsilon_pt_p + \eta_1v_1 + \cdots + \eta_qv_q\\
&\qquad = (\eta_q + \epsilon_p)t + (\epsilon_1 - \epsilon_p)t_1 + \cdots + (\epsilon_{p-1} - \epsilon_p)t_{p-1}\\
&\qquad\quad + (\eta_1 - \eta_q)v_1 + \cdots + (\eta_{q-1} - \eta_q)v_{q-1},
\end{aligned}$$

since $t_p = t - t_1 - \cdots - t_{p-1}$ and $v_q = t - v_1 - \cdots - v_{q-1}$. Hence

$$b_i = b_i^* + (\epsilon_i - \epsilon_p), \qquad (i = 1, \cdots, p-1), \tag{1}$$

and therefore

$$\begin{aligned}
\sigma_{b_ib_j} = \sigma_{b_i^*b_j^*} + \sigma_{(\epsilon_i-\epsilon_p)(\epsilon_j-\epsilon_p)} &= c_{ij}\sigma^2 + \sigma_{\epsilon_i\epsilon_j} + \sigma_{\epsilon_p\epsilon_p}\\
&= [c_{ij} + (1 + \delta_{ij})\lambda^2]\sigma^2, \quad (i, j = 1, \cdots, p-1),
\end{aligned} \tag{2}$$

where δ_{ij} is the Kronecker delta, i.e. $\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ii} = 1$. Denote $c_{ij} + (1 + \delta_{ij})\lambda^2$ by c'_{ij}. Since the expected value of b_i^* is equal to zero, on account of (1) also the expected value of b_i is equal to zero. Let

$$\| g_{ij} \| = \| c'_{ij} \|^{-1}, \qquad (i, j = 1, \cdots, p-1).$$

Then

$$\frac{1}{\sigma^2} \sum_{j=1}^{p-1} \sum_{i=1}^{p-1} g_{ij} b_i b_j \tag{3}$$

has the χ^2-distribution with $p - 1$ degrees of freedom. The expression

$$\frac{1}{\sigma^2} \sum_{\alpha=1}^{N} (y_\alpha - Y_\alpha)^2, \tag{4}$$

has the χ^2-distribution $N - p - q + 1$ degrees of freedom. The expressions (3) and (4) are independently distributed. Hence

$$\frac{N - p - q + 1}{p - 1} \frac{\Sigma\Sigma g_{ij} b_i b_j}{\Sigma (y_\alpha - Y_\alpha)^2}, \tag{5}$$

has the F-distribution (analysis of variance distribution). We will now show that (5) is a monotonic function of λ^2. It is known that $\Sigma\Sigma g_{ij}b_ib_j$ is invariant under linear transformations, i.e.

$$\Sigma\Sigma g_{ij}b_ib_j = \Sigma\Sigma g'_{ij}b'_ib'_j,$$

where b'_i is an arbitrary linear function, say $\mu_{i1}b_1 + \cdots + \mu_{ip-1}b_{p-1}$ of $b_1, \cdots, b_{p-1}$ $(i = 1, \cdots, p - 1)$ and

$$\| g'_{ij} \| = \| \sigma_{b'_ib'_j} \|^{-1}.$$

We can choose the matrix $\| \mu_{ij} \|$ such that

$$\epsilon'_i = \mu_{i1}(\epsilon_1 - \epsilon_p) + \cdots + \mu_{ip-1}(\epsilon_{p-1} - \epsilon_p), \qquad (i = 1, \cdots, p - 1),$$

are independently distributed and $\sigma^2_{\epsilon'_i} = \sigma'^2$. The coefficients μ_{ij} of course do not depend on σ'. We have

$$\sigma_{b'_ib'_j} = \sigma_{b^*_ib^*_j} + \delta_{ij}\sigma'^2, \qquad (\delta_{ij} = \text{Kronecker delta}).$$

Now let

$$b''_i = \nu_{i1}b'_1 + \cdots + \nu_{ip-1}b'_{p-1}, \qquad (\nu = 1, \cdots, p - 1),$$

where $\| \nu_{ij} \|$ is an orthogonal matrix and is chosen such that $b^{*''}_1, \cdots, b^{*''}_{p-1}$ are independently distributed. On account of the orthogonality of $\| \nu_{ij} \|$ we obviously have

$$\sigma^2_{b''_i} = \sigma^2_{b^*_i} + \sigma'^2; \qquad \sigma_{b''_ib''_j} = 0 \qquad \text{for } i \neq j.$$

Hence

$$\sum\sum g_{ij}b_ib_j = \sum_{i=1}^{p-1} \frac{b''^2_i}{\sigma^2_{b^*_i} + \lambda^2\sigma'^2}. \tag{6}$$

The right hand side of (6) is evidently a monotonic function of λ^2 which proves our statement. The endpoints of the confidence interval for λ^2 are the roots in λ^2 of the equations

$$\frac{N - p - q + 1}{p - 1} \frac{\Sigma\Sigma g_{ij}b_ib_j}{\Sigma (y_\alpha - Y_\alpha)^2} = F_2; \qquad \frac{N - p - q + 1}{p - 1} \frac{\Sigma\Sigma g_{ij}b_ib_j}{\Sigma (y_\alpha - Y_\alpha)^2} = F_1, \tag{7}$$

where F_2 denotes the upper, and F_1 the lower critical value of F

The derivation of the required confidence limits in case of classifications in more than two ways can be carried out in the same way and I shall merely state here the results.

Consider r criterions of classifications and denote by p_u the number of classes in the uth classification ($u = 1, \cdots, r$). Denote by $n_{i_1 \cdots i_r}$ the number of observations which belong to the i_1th class of the first classification, i_2th class of the second classification, $\cdots$, and to the i_rth class of the rth classification. Let $y^{(k)}_{i_1 \cdots i_r}$ be the kth observation on y in the set of observations belonging to the classes mentioned above ($k = 1, \cdots, n_{i_1 \cdots i_r}$). We make the assumption

$$y^{(k)}_{i_1 \cdots i_r} = x^{(k)}_{i_1 \cdots i_r} + \epsilon^{(1)}_{i_1} + \cdots + \epsilon^{(r)}_{i_r},$$

where the variates

$$x^{(k)}_{i_1 \cdots i_r}, \ \epsilon^{(1)}_{i_1}, \cdots, \epsilon^{(r)}_{i_r} \quad (i_u = 1, \cdots, p_u ; \ u = 1, \cdots, r; \ k = 1, \cdots, n_{i_1 \cdots i_r}),$$

are independently and normally distributed, the variance of $x^{(k)}_{i_1 \cdots i_r}$ is σ^2, the variance of $\epsilon^{(u)}_{i_u}$ is σ_u^2 and the mean value of $\epsilon^{(u)}_{i_u}$ is zero ($i_u = 1, \cdots, p_u$; $u = 1, \cdots, r$).

Let N be the total number of observations. We order the observations in a certain order and denote by y_α the αth observation in that order ($\alpha = 1, \cdots, N$). Consider the variables:

$$t, \ t^{(u)}_{i_u}, \qquad (u = 1, \cdots, r; \ i_u = 1, \cdots, p_u),$$

and denote by t_α the αth observation on t and by $t^{(u)}_{i_u}{}_\alpha$ the αth observation on $t^{(u)}_{i_u}$. The values of t_α and $t^{(u)}_{i_u \alpha}$ are given as follows:

$$t_\alpha = 1 \ (\alpha = 1, \cdots, N),$$

$t^{(u)}_{i_u \alpha} = 1$ if y_α lies in the i_uth class of the uth classification,

$t^{(u)}_{i_u \alpha} = 0$ if y_α does not lie in the i_uth class of the uth classification.

Let the sample regression of y on $t, t^{(u)}_{i_u}$ be given by

$$Y = at + \sum_{u=1}^{r} \sum_{i_u=1}^{p_u - 1} b^{(u)}_{i_u} t^{(u)}_{i_u}.$$

Let the covariance of $b^{(u)}_{i_u}$ and $b^{(u)}_{j_u}$ be given by $C^{(u)}_{i_u j_u} \sigma^2$ under the assumption that $\sigma_1 = \sigma_2 = \cdots = \sigma_r = 0$. The matrix $|| \, C^{(u)}_{i_u j_u} \, ||$ ($i_u, j_u = 1, \cdots, p_u - 1$) can be calculated by known methods of the theory of least squares. Let

$$|| \, g^{(u)}_{i_u j_u} \, || = || \, C^{(u)}_{i_u j_u} + (1 + \delta_{i_u j_u}) \lambda_u^2 \, ||^{-1} \qquad (i_u, j_u = 1, \cdots, p_u - 1),$$

where $\delta_{i_u j_u}$ is the Kronecker delta and $\lambda_u^2 = \sigma_u^2 / \sigma^2$. Then the lower and upper confidence limits for λ_u^2 are given by the roots in λ_u^2 of the equations

$$\frac{N - \sum_{u=1}^{r} p_u + r - 1}{p_u - 1} \cdot \frac{\sum_{j_u=1}^{p_u - 1} \sum_{i_u=1}^{p_u - 1} g^{(u)}_{i_u j_u} b^{(u)}_{i_u} b^{(u)}_{j_u}}{\sum_{\alpha=1}^{N} (y_\alpha - Y_\alpha)^2} = F_i \qquad (i = 1, 2), \tag{8}$$

where F_2 is the upper and F_1 the lower critical value of the analysis of variance distribution with $p_u - 1$ and $N - \sum_{u=1}^{r} p_u + r - 1$ degrees of freedom. In case of a single criterion of classification the confidence limits (8) are identical with those given in my previous paper.

THE FREQUENCY DISTRIBUTION OF A GENERAL MATCHING PROBLEM

By T. N. E. Greville

Bureau of the Census

1. Introduction. This paper considers the matching of two decks of cards of arbitrary composition, and the complete frequency distribution of correct matchings is obtained, thus solving a problem proposed by Stevens.[1] It is also shown that the results can be interpreted in terms of a contingency table.

Generalizing a problem considered by Greenwood,[2] let us consider the matching of two decks of cards consisting of t distinct kinds, all the cards of each kind being identical. The first or "call" deck will be composed of i_1 cards of the first kind, i_2 of the second, etc., such that

$$i_1 + i_2 + i_3 + \cdots + i_t = n;$$

and the second or "target" deck will contain j_1 cards of the first kind, j_2 of the second, etc., such that

$$j_1 + j_2 + \cdots + j_t = n.$$

Any of the i's or j's may be zero. It is desired to calculate, for a given arrangement of the "call" deck, the number of possible arrangements of the "target" deck which will produce exactly r matchings between them ($r = 0, 1, 2, \cdots, n$). It is clear that these frequencies are independent of the arrangement of the call deck. For convenience the call deck may be thought of as arranged so that all the cards of the first kind come first, followed by all those of the second kind, and so on.

2. Formulae for the frequencies. Let us consider the number of arrangements of the target deck which will match the cards in the k_1th, k_2th, $\cdots$, k_sth positions in the call deck, regardless of whether or not matchings occur elsewhere. Let the cards in these s positions in the call deck consist of c_1 of the first kind, c_2 of the second, etc. Then:

$$c_1 + c_2 + \cdots + c_t = s.$$

The number of such arrangements of the target deck is

$$(1) \qquad \frac{(n-s)!}{\prod_{h=1}^{t} (j_h - c_h)!}.$$

[1] W. L Stevens, *Annals of Eugenics*, Vol. 8 (1937), pp. 238–244.

[2] J. A. Greenwood, *Annals of Math. Stat*, Vol 9 (1938), pp. 56–59

For fixed values of the c's, the s specified positions may be selected in

$$\prod_{h=1}^{t} \frac{i_h!}{c_h!(i_h - c_h)!} \tag{2}$$

ways.

Consider now the expression

$$V_s = \sum \frac{(n-s)! \prod_{h=1}^{t} v_h!}{\prod_{h=1}^{t} c_h!(i_h - c_h)!(j_h - c_h)!} \tag{3}$$

obtained by summing the product of (1) and (2) over all sets of values of the numbers $c_1, c_2 \cdots, c_t$ satisfying the conditions:

$$0 \leqq c_h \leqq i_h, \qquad c_h \leqq j_h, \qquad \text{and} \qquad \sum_{h=1}^{t} c_h = s.$$

Let W_s denote the number of arrangements of the target deck which result in exactly s matchings. Then it is evident that V_s exceeds W_s, since the former includes those arrangements which give more than s matchings, and these, moreover, are counted more than once. Consider an arrangement which produces u matchings, where $u > s$. Such an arrangement will be counted once in V_s for every set of s matchings which can be selected from the total of u—that is ${}^{u}C_s$ times. In other words,

$$V_r = W_r + {}^{r+1}C_r W_{r+1} + {}^{r+2}C_r W_{r+2} + \cdots + {}^{n}C_r W_n.$$

It has been shown[3] that the solution of these equations is

$$W_r = V_r - {}^{r+1}C_r V_{r+1} + {}^{r+2}C_r V_{r+2} - \cdots + (-1)^{n-r} {}^{n}C_r V_n. \tag{4}$$

3. Computation of the frequencies. Equations (3) and (4) apparently give the solution of the problem, but in practice the labor of carrying out the summation indicated in (3) would often be very great. However, (3) may be rewritten in the form

$$V_s = \frac{(n-s)!}{\prod_{h=1}^{t} j_h!} H_s, \tag{5}$$

where

$$H_s = \sum \left\{ \prod_{h=1}^{t} \frac{i_h! \, j_h!}{c_h! \, (i_h - c_h)! \, (j_h - c_h)!} \right\}.$$

[3] H. Geiringer, *Annals of Math. Stat.*, Vol. 9 (1938), p. 262.

It will be seen that H_s is the coefficient of x^s in the product

$$\prod_{h=1}^{t}\left\{\sum_{k=0}^{i_h'}\frac{i_h!\,j_h!\,x^k}{k!\,(i_h-k)!\,(j_h-k)!}\right\}, \tag{6}$$

where i_h' denotes the smaller of i_h and j_h. The factor $\prod_{h=1}^{t} j_h!$ was included in H_s in order to make the coefficients in the polynomials of (6) always integers.

Equation (4) may now be written in the form

$$W_r = \sum_{s=r}^{n}(-1)^{s-r}\,{}^sC_r\,\frac{(n-s)!}{\prod_{h=1}^{t} j_h!}\,H_s,$$

or

$$W_r = \frac{1}{r!}\sum_{s=r}^{n}\frac{(-1)^{s-r}}{(s-r)!}\,\frac{s!\,(n-s)!}{\prod_{h=1}^{t} j_h!}\,H_s, \tag{7}$$

a form which lends itself to actual computation.

4. Factorial moments. The factorial moments of the frequency distribution of the number of matchings are easy to compute. Let m_s denote the sth factorial moment, so that

$$m_s = \frac{\sum_{r=s}^{n} r^{(s)}\,W_r}{\sum_{r=0}^{n} W_r}. \tag{8}$$

Substituting from (4)

$$\sum_{r=s}^{n} r^{(s)}W_r = \sum_{r=s}^{n}\left\{r^{(s)}\sum_{u=r}^{n}(-1)^{u-r}\,{}^uC_r\,V_u\right\}.$$

Reversing the order of summation and simplifying,

$$\sum_{r=s}^{n} r^{(s)}W_r = \sum_{u=s}^{n}\left\{u^{(s)}\,V_u\sum_{r=s}^{u}(-1)^{u-r}\,{}^{u-s}C_{r-s}\right\} = s!\,V_s.$$

Hence,

$$V_0 = \sum_{r=0}^{n} W_r = \frac{n!}{\prod_{h=1}^{t} j_h!}, \tag{9}$$

and from (5) and (8),

$$m_s = \frac{H_s}{{}_nC_s}. \tag{10}$$

5. Mean and variance. From (6)

$$H_1 = \sum_{h=1}^{t} i_h j_h \tag{11}$$

and

$$H_2 = \tfrac{1}{2} \sum_{h=1}^{t} i_h(i_h - 1) j_h(j_h - 1) + \sum_{\substack{h,k=1 \\ h \neq k}}^{t} i_h i_k j_h j_k. \tag{12}$$

Hence the mean number of matchings is

$$m_1 = \frac{\sum_{h=1}^{t} i_h j_h}{n}. \tag{13}$$

The variance μ_2 is

$$m_2 + m_1 - m_1^2 = \frac{1}{n^2(n-1)} \left[n \sum_{h=1}^{t} i_h(i_h - 1) j_h(j_h - 1) + 2n \sum_{\substack{h,k=1 \\ h<k}}^{t} i_h i_k j_h j_k \right.$$
$$\left. + n(n-1) \sum_{h=1}^{t} i_h j_h - (n-1) \left(\sum_{h=1}^{t} i_h j_h \right)^2 \right],$$

or

$$\mu_2 = \frac{1}{n^2(n-1)} \left\{ \left(\sum_{h=1}^{t} i_h j_h \right)^2 - n \sum_{h=1}^{t} i_h j_h (i_h + j_h) + n^2 \sum_{h=1}^{t} i_h j_h \right\}. \tag{14}$$

In the special case $j_1 = j_2 = \cdots = j_t = j$, these formulae become

$$M_1 = j, \qquad \mu_2 = \frac{j}{n(n-1)} \left(n^2 - \sum_{h=1}^{t} i_h^2 \right).$$

These formulae have previously been given by Stevens,[4] and those for the special case also by Greenwood. The maximal conditions for the variance, given by Greenwood for this particular case, apparently can not be put in a simple form for the general case

6. Unequal decks. Suppose the call deck contains m cards, $m < n$, and is to be matched with m cards selected from the target deck It can be assumed without loss of generality that the first m cards in any arrangement of the target deck are the ones to be used. The formulae of this paper can be applied to this

[4] W L. Stevens, *Annals of Eugenics*, loc cit , *Psychol. Review*, Vol. 46 (1939), pp. 142–150.

more general problem by the expedient of imagining $n - m$ blank cards to be added at the end of the call deck and regarding these as an additional kind. It is thus apparent that formulae (13) and (14) apply without modification to this altered situation

7. Application to contingency table. Stevens[5] has considered the distribution of entries in a contingency table with fixed marginal totals, and has pointed out that the problem of matching two decks of cards may be dealt with from that standpoint A contingency table classifies data into n columns and m rows, and we may consider the row as indicating the kind of card which occupies a given position in the call deck, the columns having the same function with respect to the target deck. Stevens defines a quantity c as the sum of entries in a prescribed set of cells, subject to the condition that no two cells of the set are in the same row or column, and mentions as unsolved the problem of the exact sampling distribution of c.

We now have at our disposal the machinery for solving this problem. Following Stevens's notation, let $a_1, a_2, \cdots, a_m$ denote the fixed row totals and $b_1, b_2, \cdots, b_n$ the fixed column totals, while x_{rs} denotes the frequency of the cell in the rth row and the sth column. Then, let $c = \sum_{h=1}^{l} x_{r_h s_h}$, where l does not exceed either m or n. Imagine two decks of N cards $\left(N = \sum_{h=1}^{m} a_h = \sum_{h=1}^{n} b_h\right)$, the first containing a_1 cards of one kind, a_2 of another, etc., and the second containing b_1 cards of one kind, b_2 of another, etc Moreover, let the r_hth kind in the first deck and the s_hth kind in the second deck be the same kind ($h = 1, 2, \cdots, l$), the other kinds being all different. Evidently c is the number of matchings between the two decks Hence, the methods of this paper can be used to obtain the distribution of c The formulae we have obtained agree with those for the expected value and variance of c given by Stevens.

ON METHODS OF SOLVING NORMAL EQUATIONS

By Paul G. Hoel

University of California, Los Angeles

There seems to be considerable disagreement concerning what is the most satisfactory method of solving a set of normal equations Since such information as errors of estimate and significance of results is usually desired in addition to the solution, in its broader aspects the problem is one of deciding what is the most satisfactory method of calculating the inverse of a symmetric matrix.

For equations with several unknowns some compact systematic method of

---

[5] W. L. Stevens, *Annals of Eugenics*, loc. cit.

calculation is necessary to eliminate much of the labor involved in the ordinary method of calculating the inverse from its definition. Among the more common of such systematic methods are those associated with the names of Chio,[1] Gauss,[1] Doolittle,[2] and Aitken.[3] In addition, A. A Albert[4] recently called attention to a method implicit in elementary matrix theory There are also various iterative schemes, and schemes which are but slight variations of the above methods In this note only the methods associated with the above names will be considered, and for convenience they will be labeled with those names, regardless of who should be given credit for them.

The purpose of this note is to show that when the calculation of the inverse is systematized, all of the above methods are fundamentally equivalent and merely involve a different arrangement of work Consequently, any advantage in calculating time for any particular method will arise through such features as a simpler technique or less copying, rather than through fewer multiplications and divisions

By the method of Chio is meant the evaluation of determinants by the pivotal method of reduction. Since all of the methods mentioned above use pivotal reduction, the method of Chio will not be treated as a distinct method Furthermore, since Gauss' method is incorporated in that of Aitken, it will be necessary to consider only the methods of Aitken, Doolittle, and Albert as distinct.

First consider the method of Albert, which is based on the following matrix properties. Let the matrix $\mathbf{A}$ be subjected to a sequence of row transformations leading to the matrix $\mathbf{A}'$. Then, writing $\mathbf{A} = \mathbf{IA}$, it follows from a theorem in matrix theory that $\mathbf{A}' = \mathbf{I}'\mathbf{A}$, and consequently that $\mathbf{A}'\mathbf{A}^{-1} = \mathbf{I}'$. If row transformations are chosen which make $\mathbf{A}' = \mathbf{I}$, then $\mathbf{A}^{-1} = \mathbf{I}'$ This states that if the same row transformations are applied to the identity matrix as were used to reduce $\mathbf{A}$ to the identity matrix, then the resulting matrix will be the desired inverse. The customary manner of reducing $\mathbf{A}$ to $\mathbf{I}$ is to work for zeros in columns as follows:

$$\begin{Vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{Vmatrix}, \qquad a_{11} \begin{Vmatrix} 1 & \frac{a_{12}}{a_{11}} & \cdots & \frac{a_{1n}}{a_{11}} \\ 0 & \left(a_{22} - a_{12}\frac{a_{21}}{a_{11}}\right) & \cdots & \left(a_{2n} - a_{1n}\frac{a_{21}}{a_{11}}\right) \\ \vdots & \vdots & & \vdots \\ 0 & \left(a_{n2} - a_{12}\frac{a_{n1}}{a_{11}}\right) & \cdots & \left(a_{nn} - a_{1n}\frac{a_{n1}}{a_{11}}\right) \end{Vmatrix},$$

[1] See, for example, Whittaker and Robinson, *The Calculus of Observations*, p 71 and p. 234

[2] See, for example, Croxton and Cowden, *Applied General Statistics*, 1939, p. 716.

[3] *Roy. Soc Edin Proc*, Vol. 57 (1936-37), p. 172

[4] Am. Math Monthly, Vol 48, No 3 (1941), p 198

$$a_{11}b_{22}\left\|\begin{array}{ccccc} 1 & \frac{a_{12}}{a_{11}} & \frac{a_{13}}{a_{11}} & \cdots & \frac{a_{1n}}{a_{11}} \\ 0 & 1 & \frac{b_{23}}{b_{22}} & \cdots & \frac{b_{2n}}{b_{22}} \\ 0 & 0 & \left(b_{33}-b_{23}\frac{b_{32}}{b_{22}}\right) & \cdots & \left(b_{3n}-b_{2n}\frac{b_{32}}{b_{22}}\right) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \left(b_{n3}-b_{23}\frac{b_{n2}}{b_{22}}\right) & \cdots & \left(b_{nn}-b_{2n}\frac{b_{n2}}{b_{22}}\right) \end{array}\right\|, \cdots,$$

where new letters are introduced for new elements after each reduction. After zeros are obtained below the main diagonal, zeros are obtained above the diagonal by starting with the last column. If now these operations are performed in the same order on I, the result will be $\mathbf{A}^{-1}$.

Next consider the method of Aitken, which is based on the evaluation of a bordered determinant, namely,

$$\left|\begin{array}{ccccccc} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} & 0 \\ \vdots & & \vdots & & \vdots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} & 1 \\ \vdots & & \vdots & & \vdots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} & 0 \\ 0 & \cdots & -1 & \cdots & 0 & 0 \end{array}\right| = \text{cofactor of } a_{ij}.$$

To obtain $\mathbf{A}^{-1}$ it is merely necessary to evaluate determinants of this type and divide them by $|A|$. Aitken's method evaluates all such determinants simultaneously, using Chio's reduction technique in much the same manner as illustrated above with Albert's method. Thus,

$$\left\|\begin{array}{cccc|cccc} a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 & 0 & \cdots & 1 \\ \hline -1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -1 & 0 & 0 & \cdots & 0 \end{array}\right\|,$$

$$
a_{11}\left\|\begin{array}{cccc|ccc}
1 & \frac{a_{12}}{a_{11}} & \cdots & \frac{a_{1n}}{a_{11}} & \frac{1}{a_{11}} & 0 \cdots & 0 \\
0 & \left(a_{22}-a_{12}\frac{a_{21}}{a_{11}}\right) & \cdots & \left(a_{2n}-a_{1n}\frac{a_{21}}{a_{11}}\right) & -\frac{a_{21}}{a_{11}} & 1 \cdots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\
0 & \left(a_{n2}-a_{12}\frac{a_{n1}}{a_{11}}\right) & \cdots & \left(a_{nn}-a_{1n}\frac{a_{n1}}{a_{11}}\right) & -\frac{a_{n1}}{a_{11}} & 0 \cdots & 1 \\
\hline
0 & \frac{a_{12}}{a_{11}} & \cdots & \frac{a_{1n}}{a_{11}} & \frac{1}{a_{11}} & 0 \cdots & 0 \\
0 & -1 & \cdots & 0 & 0 & 0 \cdots & 0 \\
\vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & 0 & \cdots & -1 & 0 & 0 \cdots & 0
\end{array}\right\|,
$$

$$
a_{11}b_{22}\left\|\begin{array}{cccc|ccc}
1 & \frac{a_{12}}{a_{11}} & \frac{a_{13}}{a_{11}} & \cdots \frac{a_{1n}}{a_{11}} & \frac{1}{a_{11}} & 0 & \cdots 0 \\
0 & 1 & \frac{b_{23}}{b_{22}} & \cdots \frac{b_{2n}}{b_{22}} & -\frac{a_{21}}{a_{11}b_{22}} & \frac{1}{b_{22}} & \cdots 0 \\
0 & 0 & \left(b_{33}-b_{23}\frac{b_{32}}{b_{22}}\right) & \cdots \left(b_{3n}-b_{2n}\frac{b_{32}}{b_{22}}\right) & \left(\frac{a_{21}}{a_{11}}\frac{b_{32}}{b_{22}}-\frac{a_{31}}{a_{11}}\right) & -\frac{b_{32}}{b_{22}} & \cdots 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \left(b_{n3}-b_{23}\frac{b_{n2}}{b_{22}}\right) & \cdots \left(b_{nn}-b_{2n}\frac{b_{n2}}{b_{22}}\right) & \left(\frac{a_{21}}{a_{11}}\frac{b_{n2}}{b_{22}}-\frac{a_{n1}}{a_{11}}\right) & -\frac{b_{2n}}{b_{22}} & \cdots 1 \\
\hline
0 & 0 & \left(\frac{a_{13}}{a_{11}}-\frac{a_{12}}{a_{11}}\frac{b_{23}}{b_{22}}\right) & \cdots \left(\frac{a_{1n}}{a_{11}}-\frac{a_{12}}{a_{11}}\frac{b_{2n}}{b_{22}}\right) & \left(\frac{a_{21}^2}{a_{11}^2b_{22}}+\frac{1}{a_{11}}\right) & -\frac{a_{12}}{a_{11}b_{22}} & \cdots 0 \\
0 & 0 & \frac{b_{23}}{b_{22}} & \cdots \frac{b_{2n}}{b_{22}} & -\frac{a_{12}}{a_{11}b_{22}} & \frac{1}{b_{22}} & \cdots 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots -1 & 0 & 0 & \cdots 0
\end{array}\right\|, \cdots.
$$

When zeros are obtained below the main diagonal to the left of the vertical dividing line, the matrix in the lower right section will be $\mathbf{A}^{-1}$. This follows from the fact that the elements of this matrix will be the evaluations of bordered determinants, like those of the previous paragraph, divided by $a_{11}b_{22}\cdots = |A|$

It will be observed that the operations on **A** in Albert's method which produce zeros below the main diagonal are the same as those which occur above the horizontal dividing line in Aitken's method. This set of operations is performed simultaneously on **I**, since the upper right section of Aitken's scheme is I Furthermore, obtaining a zero for an element below the horizontal line and to the left of the vertical line, is equivalent to obtaining a zero for the element corre-

sponding to the same row and column in the section above the horizontal, provided the preceding columns contain zeros above the diagonal. But obtaining zeros above the main diagonal of A constitutes the second set of operations in Albert's method to obtain $A' = I$. Thus, the operations in Aitken's method which produce zeros in a given column for elements above the horizontal line are merely the first set of operations in Albert's method, while those which produce zeros below the horizontal line are the second set of operations in reverse order. Since, in Aitken's scheme, the first set of operations is performed on I in the upper right section and the results are transferred a row at a time to the lower right section, where they are in turn operated upon by the second set of operations, this lower right section is merely I operated upon by the entire set of operations of Albert's method. Consequently, Aitken's and Albert's methods are the same except for the order in which operations are performed and differences arising therefrom. Since Aitken's method performs these operations more compactly, it is to be preferred to that of Albert.

Next consider the method of Doolittle, which is described by following the instructions given in the first column in the table shown on page 318. The forward solution is completed after n such sectional operations. For a given k column, the backward solution is obtained as usual by substitution in the last row of each section taken in reverse order.

If all summations in each section are performed in pairs and the sums recorded each time, rather than being performed in one operation, the forward solution of the Doolittle method will be found to be a rearrangement of the work occurring above the horizontal line in Aitken's method. Thus the first lines of each section give the matrix above the horizontal line in Aitken's scheme. Then, except for signs, I' and the sums of the first two lines of the remaining sections give the result of Aitken's first sequence of operations above the horizontal. Then, except for signs, II' and the sums of the first three lines of the remaining sections give the result of Aitken's second sequence of operations above the horizontal, etc.

The back solution involves precisely the same operations as those making up the second set of Albert's sequence of operations to obtain zeros above the main diagonal. Since these were shown to be a rearrangement of operations in Aitken's method, it follows that the methods of Aitken and Doolittle are the same except for the order of operations and differences arising therefrom. Hence all three methods are basically the same when systematized for a calculating machine.

Because of this equivalence, the number of necessary multiplications and divisions will be the same for all three methods, and will be found to be $\frac{1}{2}n^2(n + 1)$. Since Aitken's method is to be preferred to that of Albert, it will suffice to compare the methods of Aitken and Doolittle for calculating convenience.

The Doolittle method possesses several distinct advantages. First, its multiplications occur a row at a time with one of the factors constant for that row; consequently the keyboard remains unchanged for a given row of operations.

	1	2	3		n	k_1	k_2	..	k_n
I and ΣI	a_{11}	a_{12}	a_{13}		a_{1n}	-1	0		0
I′	-1	$-\frac{a_{12}}{a_{11}}$	$-\frac{a_{13}}{a_{11}}$	..	$-\frac{a_{1n}}{a_{11}}$	$\frac{1}{a_{11}}$	0		0
II	a_{21}	a_{22}	a_{23}	..	a_{2n}	0	-1		0
ΣI I′$_2$	$-a_{12}$	$-a_{12}\frac{a_{12}}{a_{11}}$	$-a_{13}\frac{a_{12}}{a_{11}}$	...	$-a_{1n}\frac{a_{12}}{a_{11}}$	$\frac{a_{12}}{a_{11}}$	0		0
ΣII	0	$\left(a_{22}-a_{12}\frac{a_{12}}{a_{11}}\right)$	$\left(a_{23}-a_{13}\frac{a_{12}}{a_{11}}\right)$		$\left(a_{2n}-a_{1n}\frac{a_{12}}{a_{11}}\right)$	$\frac{a_{12}}{a_{11}}$	-1	...	0
II′	0	-1	$-\frac{b_{23}}{b_{22}}$		$-\frac{b_{2n}}{b_{22}}$	$-\frac{a_{12}}{a_{11}b_{22}}$	$\frac{1}{b_{22}}$		0
III	a_{31}	a_{32}	a_{33}	..	a_{3n}	0	0		0
ΣI·I′$_3$	$-a_{13}$	$-a_{12}\frac{a_{13}}{a_{11}}$	$-a_{13}\frac{a_{13}}{a_{11}}$	..	$-a_{1n}\frac{a_{13}}{a_{11}}$	$\frac{a_{13}}{a_{11}}$	0		0
ΣII·II′$_3$	0	$-b_{23}$	$-b_{23}\frac{b_{23}}{b_{22}}$	..	$-b_{2n}\frac{b_{23}}{b_{22}}$	$-\frac{a_{12}}{a_{11}}\frac{b_{23}}{b_{22}}$	$\frac{b_{23}}{b_{22}}$	..	0
ΣIII	0	0	c_{33}	...	c_{3n}	c_{3n+1}	$\frac{b_{23}}{b_{22}}$		0
III′	0	0	-1	.	$-\frac{c_{3n}}{c_{33}}$	$-\frac{c_{3n+1}}{c_{33}}$	$-\frac{b_{23}}{b_{22}c_{33}}$		0
⋮			⋮				⋮	⋮	⋮

Aitken's method, however, consists of calculating successive cross products' which requires clearing of the keyboard after each such operation. Secondly, there are fewer additions in the Doolittle method. It sums z quantities at a time in section z, while Aitken's cross products always involve the sum of two quantities. Because of the necessity of calculating the complements of negative sums, this difference becomes important when the number of variables is large. A third feature in favor of the Doolittle method is the ease of performing the calculations without previous experience It may be easier to understand how to calculate cross products, but actually the calculations of the Doolittle method are easier to perform. Aitken's method requires some experience with it, if one is to avoid repeating certain calculations which would result from calculating all cross products mechanically. The comparative amount of copying in the two methods depends upon the number of variables involved.

From the above considerations, it may be concluded that the Doolittle method is to be preferred among those considered in this paper for solving a set of normal equations or calculating the inverse of a symmetric matrix However, if a single calculating technique is desired which can be used for nonsymmetrical equations as well, then the method of Aitken is to be preferred.

CONDITIONS THAT THE ROOTS OF A POLYNOMIAL BE LESS THAN UNITY IN ABSOLUTE VALUE

By Paul A. Samuelson

Massachusetts Institute of Technology

1. Introduction. In econometric business cycle analysis, probability theory, and numerical mathematical computation the problem of convergence of repeated iterations arises. The solution of the difference equations defining such a process can in a wide variety of cases be shown to be stable in the sense of converging to a limit if a certain associated polynomial

$$f(x) = p_0x^n + p_1x^{n-1} + \cdots + p_n = 0, \tag{1}$$

has roots whose moduli are all less than unity.

Thus, for "timeless" linear difference equation systems of the most general type, convertible into normal form,

$$Q_i(t+1) = \sum_{j=1}^{n} a_{ij}Q_j(t), \qquad (i = 1, \cdots, n), \tag{2}$$

the polynomial is the characteristic or determinantal equation,

$$f(x) = |\, a_{ij} - x\delta_{ij} \,| = 0, \tag{3}$$

which when expanded out is of the form (1). The roots of this equation, when multiplied by suitable polynomials in t, give the exact solution of the problem in the form

$$Q(t) = \sum_{i=1}^{m} g_i(t)x_i^t, \tag{4}$$

where m is the number of distinct roots, and the g's are polynomials of degree one less than the multiplicity of the respective root. If complex roots occur, they do so in conjugate pairs and can be combined to form damped, undamped, or anti-damped harmonic terms All terms go to zero as t approaches infinity if, and only if, the absolute value of each x is less than unity.

For non-linear systems the exact solution does not take this form, but in the neighborhood of an equilibrium point the roots of an associated polynomial, except in singular cases, do determine the stability of the system.

As far as the writer is aware, there does not appear in the literature an account of necessary and sufficient conditions for the roots of a polynomial to be less than unity in absolute value. This is in contrast to a related problem which arises in connection with the investigation of stability of dynamical systems defined by differential equations. These have associated with them a polynomial whose roots provide solutions in the form

$$g_i(t)e^{x_it}, \tag{5}$$

or for non-linear systems infinite power series in such terms It is required, therefore, to determine complete conditions under which the *real parts* of all roots must be negative

This problem has been solved by Routh[1] in a manner which leaves little to be desired. Determinantal expression of his conditions in a slightly modified form was made by Hurwitz[2] who apparently was unaware of Routh's work, and by Frazer and Duncan[3] who were unaware of the Hurwitz results. A brief outline of Routh's mode of attack will prove instructive in dealing with the problem at hand.

2. Routhian analysis of sign of real parts of roots. Routh realized that the condition that all coefficients be positive—the leading coefficient having been made so—was necessary, but not sufficient unless all the roots were real. But a "derived" equation of degree $n(n-1)/2$ whose roots equal the sums of the roots of the original equation taken two at a time has real roots which are simple sums of the real parts of those of the original equation. In consequence, it is necessary and sufficient that the coefficients of the original and the "derived" equation all be positive.

Thus, valid necessary and sufficient conditions are presented. However, they are disadvantageous from two points of view. First, they are not all independent, being $n(n+1)/2$ conditions in number, whereas only n are necessary. Secondly, despite several ingenious methods devised by Routh, it is not easy to compute them in the general case.

Recognizing these difficulties, he therefore began anew from an entirely different angle Utilizing a theorem of Cauchy concerning the relationship between the behavior of a polynomial on a closed contour in the complex domain and the number of roots within that closed curve, he derived necessary and sufficient conditions, which may be written in the slightly more convenient determinantal form of Hurwitz and Frazer and Duncan as follows.

$$T_0 = p_0 > 0, \qquad T_1 = p_1 > 0, \qquad T_2 = \begin{vmatrix} p_1 & p_3 \\ p_0 & p_2 \end{vmatrix} > 0,$$

$$(6) \qquad T_3 = \begin{vmatrix} p_1 & p_3 & p_5 \\ p_0 & p_2 & p_4 \\ 0 & p_1 & p_3 \end{vmatrix} > 0, \quad \cdots \quad T_s = \begin{vmatrix} p_1 & p_3 & \cdots & p_{2s-1} \\ p_0 & p_2 & \cdots & p_{2s-2} \\ 0 & p_1 & \cdots & p_{2s-3} \\ 0 & p_0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & p_s \end{vmatrix} > 0.$$

[1] E. J. Routh, *A Treatise on the Stability of a Given State of Motion*, (London, 1877), Chaps. 2 and 3, *Advanced Rigid Dynamics*, 6th ed , London, 1905, Chap. 6

[2] Hurwitz, *Math Ann* , Vol. 46 (1895), p 521

[3] R A. Frazer and W. J. Duncan, *Royal Soc Proc* , Series A, Vol 124 (1929), p. 642. Also R A Frazer, W J Duncan, and A R. Collar, *Elementary Matrices*, Cambridge University Press, 1938, pp 151–155.

The law of formation of these determinants is obvious. In the first row the odd p's starting with the first are listed. Within each column the p's diminish one unit at a time. Any p with negative subscript derived by this formula is treated as zero, and all p's of subscript higher than the degree of the equation are set equal to zero. With this convention, for p_0 made positive, complete and independent necessary conditions are that all principal minors of T_n formed by deleting successively the last row and column must be positive. These conditions are n in number and are independent.

3. Complete, independent, necessary and sufficient conditions. Corresponding to Routh's first attack on the problem, we might consider an equation of degree $n(n-1)/2$ whose roots equal the *products* two at a time of the original equation's. If this equation and the original equation have *real* roots less than unity in absolute value, our problem is solved. This is guaranteed if, and only if, two further transformed equations with roots equal to the squares minus unity of the roots of the original and derived equations respectively all have positive coefficients. These conditions are necessary and sufficient, but not independent, and cannot be easily computed in the general case. Therefore, I follow Routh's example and approach the problem from a different point of view.

When the roots of $f(x) = 0$ are plotted in the complex plane, they must all lie within the unit circle if their absolute values are to be less than unity, and conversely. We might therefore attempt to apply Cauchy's theorem. However, it is not necessary to do so. Routh has shown what the conditions are that there be no roots in the right-hand half-plane. Can we find a complex transformation of variables which carries the unit circle into the left-hand half-plane?

The answer is in the affirmative. The linear complex transformation

$$x = \frac{z+1}{z-1}, \qquad z = \frac{x+1}{x-1} \tag{7}$$

will accomplish this. But after substituting for x its value in terms of z, we cease to have a polynomial but rather a rational function of z as follows:

$$f(x) = f\left(\frac{z+1}{z-1}\right) = \frac{\sum_{i=0}^{n} p_i (z+1)^{n-i}(z-1)^i}{(z-1)^n} = 0. \tag{8}$$

We need only consider the polynomial in the numerator, i.e.,

$$\varphi(z) = \sum_{0}^{n} \pi_i z^{n-i} = 0. \tag{9}$$

In order that the roots of the original equation be less than unity, in absolute value, it is necessary and sufficient that the real parts of the roots of equation (9) *be negative.* Once we determine the coefficients (π_i) in terms of the original p's, we can easily apply Routh's theorems. This yields $n+1$ necessary and sufficient conditions, all of which are independent.

Expanding the numerator of the right-hand side of (8) and collecting terms, the following explicit formulas for the π's are directly obtained:

$$\pi_i = \sum_{j=0}^{n} p_j \sum_{k=0}^{m(i,j)} {}_{n-j}C_{i-k}(-1)^k {}_jC_k, \tag{10}$$

where

$${}_vC_w = \frac{v!}{(v-w)!\,w!},$$

and

$m(i, j)$ = the smaller of i and j.

For fourth and higher degree equations literal substitution, while always possible, results in complicated expressions It is preferable, therefore, to compute the π's numerically and then apply the conditions of (6) directly.

Other necessary conditions can be easily derived, but they will be dependent upon these. Thus, each π must be positive; but this is not, by itself, sufficient. Or, adding π_0 and π_n we find

$$\pi_0 + \pi_n = p_0 + p_2 + p_4 + \cdots > 0, \tag{11}$$

i e., the sum of the even p's must be positive. Similarly, still other linear sums of other π's will result in cancellation of certain of the p's. Except on special occasions there is probably no labor saved by utilizing conditions derived in this way.

One obvious but useful necessary condition will be stated without proof. If one forms polynomials from subsets of the coefficients of a given "stable" polynomial formed by arbitrary "cuts" which leave adjacent coefficients in unchanged order and introduce no gaps within each set, then the resulting polynomials will all be stable.

Special sufficiency conditions also can be developed. Carmichael[4] presents certain inequalities between the absolute values of the largest root and the coefficients of the original equation. For special problems these may be fruitfully applied.

4. Example. In conclusion I apply the conditions derived here to a well-known numerical equation determined statistically by Tinbergen[5] in the analysis of economic fluctuations. It is a fourth order difference equation with constant coefficients,

$$Z_t - 398Z_{t-1} + .220Z_{t-2} - .013Z_{t-3} - .027Z_{t-4} = 0 \tag{12}$$

[4] R D. Carmichael, *Amer Math Soc. Bull*, Vol 24 (1918), pp. 286–296.

[5] J. Tinbergen, *Business Cycles in the United States, 1919–1932*, League of Nations, 1939, p. 140.

with the associated indicial equation

$$f(x) = x^4 - .398x^3 + .220x^2 - .013x - .027 = 0. \tag{13}$$

Its roots have been computed and are known to be less than unity in absolute value. This may be verified by computing

$$\begin{aligned} \pi_0 &= 0.782 > 0 \\ \pi_1 &= 3.338 > 0 \\ \pi_2 &= 5.398 > 0 \\ \pi_3 &= 4.878 > 0 \\ \pi_4 &= 1.604 > 0 \\ T_2 &= 14.204 > 0 \\ T_3 &= 43.177 > 0 \end{aligned} \tag{14}$$

To compute the same results by cross-multiplication the work is arranged as follows:

$$\begin{array}{ccc} \pi_0 & & \pi_2 \quad\quad \pi_4 \\ .782 & & 5.398 \quad\quad 1.604 \\ \pi_1 & & \pi_3 \\ 3.338 & & 4.878 \\ \pi_1\pi_2 - \pi_0\pi_3 & & \pi_3\pi_4 - 0 \\ 14.204 & & 7.824 \\ \pi_3(\pi_1\pi_2 - \pi_0\pi_3) - \pi_1\pi_3\pi_4 & & \\ 43.177 & & \end{array} \tag{15}$$

It may be remarked that the presence of a negative coefficient anywhere in the table is an immediate indication of instability, and that there is no necessity to continue the computation until a negative sign appears in a leading coefficient. This fact often saves much labor.

VALUES OF MILLS' RATIO OF AREA TO BOUNDING ORDINATE AND OF THE NORMAL PROBABILITY INTEGRAL FOR LARGE VALUES OF THE ARGUMENT

By Robert D. Gordon

Scripps Institution of Oceanography

A pair of simple inequalities is proved which constitute upper and lower bounds for the ratio R_x^{-1}, valid for $x > 0$. The writer has failed to encounter these inequalities in the literature, hence it seems worthwhile to present them for whatever value they may have.

[1] J. P. Mills, "Table of ratio: area to bounding ordinate, for any portion of the normal curve." *Biometrika Vol. 18* (1926) pp. 395–400. Also Pearson's tables, Part II, Table III.

The function R_x is defined by

$$R_x = e^{x^2/2} \int_x^\infty e^{-t^2/2}\, dt. \tag{1}$$

The following relations between $R = R_x$ and its derivatives are easily established by direct differentiations and substitutions:

$$\frac{dR}{dx} = xR - 1, \tag{2}$$

$$\frac{d^2R}{dx^2} = x\frac{dR}{dx} + R = \frac{x^2+1}{x}\frac{dR}{dx} + \frac{1}{x}, \tag{3}$$

$$\frac{d^3R}{dx^3} = \left(1 + \frac{2}{x^2+1}\right)x\frac{d^2R}{dx^2} - \frac{2}{x^2+1}. \tag{4}$$

Also by ordinary rules

$$R_x > 0, \tag{5}$$

$$\lim_{x\to\infty} xR_x = 1. \tag{6}$$

1°. Suppose that at any point $x_1 > 0$, $x_1R > 1$. Then by (2) $dR/dx > 0$, and R_x would continue to increase with increasing x: still more, xR_x would continue to increase, hence we should have $xR_x > 1$ for $x \geqq x_1$, which contradicts (6). Therefore we find $xR_x \leqq 1$ for $x > 0$, and

$$R_x \leqq \frac{1}{x}, \tag{7}$$

which establishes the required upper inequality.

2° Suppose that at any point $x_2 > 0$, $d^2R/dx^2 < 0$. Then by (4) $d^3R/dx^3 = (d/dx)(d^2R/dx^2) < 0$ at this point Since these derivatives are continuous this implies that for all $x > x_2$, $d^2R/dx^2 < [d^2R/dx^2]_{x=x_2} < 0$. Then we get the inequalities, for $x > x_2$

$$\frac{dR}{dx} < \left[\frac{dR}{dx}\right]_2 + (x - x_2)\left[\frac{d^2R}{dx^2}\right]_2 < \left[\frac{dR}{dx}\right]_2$$

$$R < R_{x_2} + (x - x_2)\left[\frac{dR}{dx}\right]_2 + \tfrac{1}{2}(x - x_2)^2\left[\frac{d^2R}{dx^2}\right]_2$$

where $[\ ]_2$ indicates evaluation at $x = x_2$. Since $[d^2R/dx^2]_2 < 0$, this implies that for sufficiently large x, $R_x < 0$, which contradicts (5). It follows then that (3) is positive, and substitution of (2) gives

$$R_x \geqq \frac{x}{x^2+1}. \tag{8}$$

We combine (7) and (8) in the double inequality:

$$\frac{x}{x^2+1} \leqq R_x \leqq \frac{1}{x}, \qquad \text{if } x \geqq 0. \tag{9}$$

This gives for the probability integral the corresponding inequality

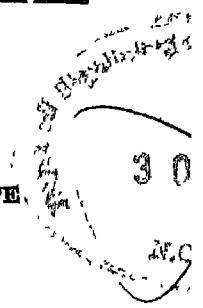
$$\frac{x}{x^2+1} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \leqq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt \leqq \frac{1}{x} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{10}$$

It can easily be shown (for $x > 0$) that equalities in (9) and (10) are impossible.

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS



Contents

PAGE

Distribution of the Ratio of the Mean Square Successive Difference to the Variance. JOHN VON NEUMANN 367

Some Examples of Asymptotically Most Powerful Tests. ABRAHAM WALD 396

On the Distribution of the Quotient of Two Chance Variables. J. H. CURTISS 409

Some Generalizations of the Logarithmic Mean and of Similar Means of Two Variates Which Become Indeterminate When the Two Variates are Equal. EDWARD L. DODD 422

A Study of R. A. Fisher's z-Distribution and the Related F-Distribution. LEO A. AROIAN 429

The Doolittle Technique. PAUL S. DWYER 449

Notes:

A Problem in Estimation. JOSEPH F. DALY 459

Confidence Limits for an Unknown Distribution Function. A. KOLMOGOROFF 461

Corrections to a Paper on the Uniqueness Problem of Moments. M. G. KENDALL 464

Announcement Concerning Computation of Mathematical Tables 465

Report of the Chicago Meeting of the Institute 468

Abstracts of Papers 470

Vol. XII, No. 4 — December, 1941

DISTRIBUTION OF THE RATIO OF THE MEAN SQUARE SUCCESSIVE DIFFERENCE TO THE VARIANCE

By John von Neumann

Institute for Advanced Study[1]

1. Introduction. Let $x_1, \cdots, x_n$ be variables representing n successive observations in a population which obeys a distribution law

$$ce^{-(x-\xi)^2/2\sigma^2}\,dx, \qquad \left(c = \frac{1}{\sigma\sqrt{2\pi}}\right),$$

i.e. which is normal, with the mean ξ and the standard deviation σ. For the sample we define as usual the mean,

$$\bar{x} = \frac{1}{n}\sum_{\mu=1}^{n} x_\mu,$$

the variance,

$$s^2 = \frac{1}{n}\sum_{\mu=1}^{n} (x_\mu - \bar{x})^2,$$

and also the mean square successive difference

$$\delta^2 = \frac{1}{n-1}\sum_{\mu=1}^{n-1} (x_{\mu+1} - x_\mu)^2.$$

The reasons for the study of the distribution of the mean square successive difference δ^2, in itself as well as in its relationship to the variance s^2, have been set forth in a previous publication[2], to which the reader is referred. The distribution of δ^2, and in particular its moments, were also studied there. The present paper is devoted to the investigation of the ratio

$$\eta = \frac{\delta^2}{s^2}.$$

A comparison of the observed value of η with that distribution is particularly suited as a basis of the judgment whether the observations $x_1, \cdots, x_n$ are independent or whether a trend exists. (Cf. sections 1 and 2, loc. cit.[2])

The moments of η have already been determined by J. D Williams by a

---

[1] Also Scientific Advisory Committee of the Ballistic Research Laboratory, Aberdeen Proving Ground.

[2] John von Neumann, R H Kent, H R. Bellinson, B I. Hart, "The mean square successive difference," *Annals of Math Stat.*, Vol. 12 (1941), pp 153-162.

different method.[3] Williams' results have been checked by W. J. Dixon at the suggestion of S. S. Wilks, whose stimulating interest has been largely responsible for the undertaking of the series of papers on δ^2 and $\frac{\delta^2}{s^2}$. The present rather exhaustive discussion, however, brings out several other essential characteristics of this statistic, and provides the key to some very effective computational methods. It is further hoped that the reader will find that the mathematical methods used and the generalizations indicated have an interest of their own.

From the latter point of view the final results of sections 5 and 7, concerning the distribution of values of quadratic and of Hermitian forms, may deserve special attention.

2. Diagonalization of the quadratic forms and replacement by a spherical mean. Since δ^2 and s^2 are unchanged when we replace each x_μ by $x_\mu - \xi$, we may assume $\xi = 0$. Then the distribution law of x is

$$cc^{-x^2/2\sigma^2}\,dx, \qquad \text{and that of } x_1, \cdots, x_n \text{ is } \prod_{\mu=1}^{n} cc^{-x_\mu^2/2\sigma^2}\,dx_\mu,$$

i.e.

$$c^n e^{-\sum_{\mu=1}^{n} x_\mu^2/2\sigma^2}\,dx_1 \cdots dx_n.$$

Any linear orthogonal transformation of the $x_1, \cdots, x_n$ leaves $\sum_{\mu=1}^{n} x_\mu^2$ and $dx_1 \cdots dx_n$ unchanged, hence the above distribution law will likewise be left unchanged. Thus, we may subject the two quadratic forms δ^2, s^2 to any simultaneous linear, orthogonal transformation.

Consider one carrying $x_1, \cdots, x_n$ into, say $x_1', \cdots, x_n'$, which brings the quadratic form $(n-1)\delta^2$ into the diagonal form, say $\sum_{\mu=1}^{n} A_\mu x_\mu'^2$. Such a transformation does not affect the characteristic values of the quadratic forms[4], and these characteristic values are obviously $A_1, \cdots, A_n$ in the case of $\sum_{\mu=1}^{n} A_\mu x_\mu'^2$. Consequently $A_1, \cdots, A_n$ are the characteristic values of the original quadratic form $(n-1)\delta^2$. We shall determine them as such in the next section.

Clearly we always have $(n-1)\delta^2 \geqq 0$, hence all $A_\mu \geqq 0$. Some A_μ may

---

[3] J. D. Williams, "Moments of the ratio of the mean square successive difference to the mean square difference in samples from a normal universe," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 239–241. Cf. also L. C. Young, "On randomness in orded sequences," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 293–300.

[4] For the properties of matrices and quadratic forms cf. e.g.: J. H. M. Wedderburn, Lectures on Matrices, *Amer. Math. Soc. Colloquium Publications*, Vol. 17, New York, 1934. In the present context cf. mainly Chapters II and VI.

equal 0 say k $(= 0, 1, \cdots, n)$ of them, which we can arrange to be A_{n-k+1}, $\cdots$, A_n.

$(n-1)\delta^2 = 0$ is thus equivalent to $x_1' = \cdots = x_{n-k}' = 0$, i.e. to $n-k$ independent conditions. On the other hand this amounts obviously to $x_1 = \cdots = x_n$, and these are $n-1$ independent conditions. So $k = 1$ and consequently $A_1, \cdots, A_{n-1} > 0$, $A_n = 0$. And our linear orthogonal transformation must carry the x-vectors with $x_1 = \cdots = x_n$ into the x'-vectors with $x_1' = \cdots = x_{n-1}' = 0$. Among the former, $\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}$ has the length 1; among the latter only $0, \cdots, 0, \pm 1$ have. Hence these correspond to each other. Now the scalar (inner) product of two vectors is an orthogonal invariant, that of a vector $x_1, \cdots, x_n$ with $\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}$ is $\sqrt{n}\bar{x}$, that of a vector $x_1', \cdots, x_n'$ with $0, \cdots, 0, \pm 1$ is $\pm x_n'$, hence

$$\sqrt{n}\,\bar{x} = \pm x_n'.$$

Put $x_\mu = \bar{x} + u_\mu$. Then clearly $\sum_{\mu=1}^{n} u_\mu = 0$. Hence

$$\sum_{\mu=1}^{n} x_\mu^2 = n\bar{x}^2 + \sum_{\mu=1}^{n} u_\mu^2 = x_n'^2 + ns^2.$$

Owing to the orthogonality, the left-hand side is equal to $\sum_{\mu=1}^{n} x_\mu'^2$, therefore

$$ns^2 = \sum_{\mu=1}^{n-1} x_\mu'^2.$$

Remembering that $A_n = 0$, we also have

$$(n-1)\delta^2 = \sum_{\mu=1}^{n-1} A_\mu x_\mu'^2.$$

Consequently

$$\eta = \frac{\delta^2}{s^2} = \frac{n}{n-1} \frac{\sum_{\mu=1}^{n-1} A_\mu x_\mu'^2}{\sum_{\mu=1}^{n-1} x_\mu'^2}.$$

The distribution law is, as we know, the same in $x_1', \cdots, x_n'$ as in $x_1, \cdots, x_n$, namely

$$c^n e^{-\sum_1^n x_\mu'^2/2\sigma^2}\, dx_1' \cdots dx_n'.$$

Thus $x_1', \cdots, x_n'$ are independent. η depends on $x_1', \cdots, x_{n-1}'$ only, hence we may disregard x_n' altogether, and use the distribution law of the $x_1', \cdots, x_{n-1}'$,

$$c^{n-1} e^{-\sum_1^{n-1} x_\mu'^2/2\sigma^2}\, dx_1' \cdots dx_{n-1}'.$$

With respect to $x_1', \cdots, x_{n-1}'$ we may now state that the $x_1', \cdots, x_{n-1}'$ distribution of η can be obtained by determining first the distribution of η over every spherical surface

$$\sum_{\mu=1}^{n-1} x_\mu'^2 = r^2$$

and then averaging these distributions with the weights $\psi(r)\,dr$, where $\psi(r)\,dr$ is the probability of the spherical shell from r to $r + dr$ with respect to our original $x_1', \cdots, x_{n-1}'$ distribution law. (It happens to be $c'e^{-r^2/2\sigma^2} r^{n-2}\,dr$, but this is immaterial.)

Since the $x_1', \cdots, x_{n-1}'$ distribution law is obviously spherically symmetric in these variables, the first-mentioned distributions over the spherical surfaces are readily obtained by assigning each piece of the surfaces in question its own relative, $n-2$-dimensional area as weight.

Since η is a homogeneous function of $x_1', \cdots, x_{n-1}'$ of order zero, these spherical surface distributions of η are the same for all r. Consequently we can replace all these r by, say $r = 1$, and the subsequent averaging over the r may be omitted altogether.

Finally, since we restrict ourselves to $r = 1$, i.e. to the spherical surface

$$\sum_{\mu=1}^{n-1} x_\mu^2 = 1,$$

the denominator of η may be omitted and we have

$$\eta = \frac{n}{n-1} \sum_{\mu=1}^{n-1} A_\mu x_\mu'^2.$$

We sum up, writing again $x_1, \cdots, x_{n-1}$ for $x_1', \cdots, x_{n-1}'$, then the desired distribution of η is that of

$$\eta = \frac{n}{n-1} \sum_{\mu=1}^{n-1} A_\mu x_\mu^2,$$

where the point $x_1, \cdots, x_{n-1}$ is uniformly distributed over the spherical surface

$$\sum_{\mu=1}^{n-1} x_\mu^2 = 1.$$

Here $A_1, \cdots, A_{n-1}$ are all positive, and together with 0 they are the characteristic values of the quadratic form

$$(n-1)\delta^2 = \sum_{\mu=1}^{n-1} (x_{\mu+1} - x_\mu)^2$$

$$= x_1^2 + 2\sum_{\mu=2}^{n-1} x_\mu^2 + x_n^2 - 2\sum_{\mu=1}^{n-1} x_\mu x_{\mu+1}.$$

3. The characteristic values A_μ ; first orientation concerning η. We have shown that there exist (counting multiplicities) precisely $n-1$ positive roots A of the characteristic equation

$$Det \begin{vmatrix} A-1 & 1 & & & & & & & \\ 1 & A-2 & 1 & & & & & & \\ & 1 & A-2 & 1 & & & & & \\ & & 1 & A-2 & 1 & & & & \\ & & & 1 & \ddots & & & & \\ & & & & \ddots & \ddots & & & \\ & & & & & \ddots & 1 & & \\ & & & & & 1 & A-2 & 1 & \\ & & & & & & 1 & A-2 & 1 \\ & & & & & & & 1 & A-1 \end{vmatrix} = 0$$

(the empty places are filled with zeros), and that these roots are the $A_1, \cdots, A_{n-1}$.

Such an A is characterized by the possibility of solving the equations

$$(A-1)x_1 + x_2 = 0, \quad x_1 + (A-2)x_2 + x_3 = 0, \quad x_2 + (A-2)x_3 + x_4 = 0,$$

$$\cdots\cdots\cdots, \quad x_{n-2} + (A-2)x_{n-1} + x_n = 0, \quad x_{n-1} + (A-1)x_n = 0,$$

in $x_1, \cdots, x_n$ not all equal to zero. Put

$$x_0 = x_1, \qquad x_{n+1} = x_n,$$

and

$$A = 2 - 2\cos\alpha,$$

then these equations become

$$x_{\mu-1} + x_{\mu+1} = 2\cos\alpha \cdot x_\mu \qquad \text{for} \qquad \mu = 1, 2, \cdots, n-1, n.$$

The last equation is satisfied by

$$x_\mu = 2\cos(\mu - \tfrac{1}{2})\alpha \qquad \text{for} \qquad \mu = 0, 1, 2, \cdots, n-1, n, n+1.$$

Now $x_0 = x_1$ is automatically fulfilled, while $x_{n+1} = x_n$ demands $\cos(n+\tfrac{1}{2})\alpha = \cos(n-\tfrac{1}{2})\alpha$. This is certainly the case when $(n+\tfrac{1}{2})\alpha = 2k\pi - (n-\tfrac{1}{2})\alpha$ (k any integer), i.e. $\alpha = \dfrac{k\pi}{n}$. For no $k = 1, \cdots, n-1$ are $x_1, \cdots, x_n$ all equal to zero $\left(\text{indeed } x_1 = 2\cos\dfrac{k\pi}{2n} > 0\right)$, hence these k give A's of the desired kind. They are

$$A = 2 - 2\cos\frac{k\pi}{n} = 4\sin^2\frac{k\pi}{2n} \qquad (k = 1, \cdots, n-1),$$

and so they are all positive and different from each other. Their number is $n-1$. Hence they are precisely $A_1, \cdots, A_{n-1}$.

So we have shown

$$A_\mu = 2 - 2\cos\frac{\mu\pi}{n} = 4\sin^2\frac{\mu\pi}{2n} \qquad (\mu = 1, \cdots, n-1).$$

We can now reformulate the final result of the preceding section. Let us set

$$\eta = \frac{2n}{n-1}(1-\epsilon).$$

Then

$$\epsilon = \sum_{\mu=1}^{n-1} \cos\frac{\mu\pi}{n}\, x_\mu^2,$$

where the point $x_1, \cdots, x_{n-1}$ is uniformly distributed over the spherical surface

$$\sum_{\mu=1}^{n-1} x_\mu^2 = 1.$$

Replacement of x_μ by $x_{n-\mu}$ carries ϵ into $-\epsilon$. Therefore the distribution of ϵ is symmetric around 0. Hence the mean of ϵ is 0. The maximum of ϵ's distribution is clearly $\cos\frac{\pi}{n}$, its minimum is $\cos\frac{(n-1)\pi}{n} = -\cos\frac{\pi}{n}$. We state these facts, together with their equivalents for η.

ϵ (η)'s distribution is symmetric around its mean, which is 0 $\left(\frac{2n}{n-1}\right)$. The maximum of ϵ (η)'s distribution is $\cos\frac{\pi}{n}\left(\frac{2n}{n-1}\left[1+\cos\frac{\pi}{n}\right] = \frac{4n}{n-1}\cos^2\frac{\pi}{2n}\right)$, its minimum is $-\cos\frac{\pi}{n}\left(\frac{2n}{n-1}\left[1-\cos\frac{\pi}{n}\right] = \frac{4n}{n-1}\sin^2\frac{\pi}{2n}\right)$.

Thus it will be easier to obtain information concerning η by considering the distribution of ϵ, since all odd moments of ϵ are zero, etc. The investigation of ϵ instead of η was first suggested by B. I. Hart, who also found, that the first four odd moments of ϵ vanish. R. H. Kent and B. I. Hart also determined the minima and maxima of these distributions for certain small values of n.

4. Direct computation of the moments. We shall investigate the distribution law of a quantity

$$\gamma = \sum_{\mu=1}^{m} B_\mu x_\mu^2.$$

where the point $x_1, \cdots, x_m$ is equidistributed over the spherical surface

$$\sum_{\mu=1}^{m} x_\mu^2 = 1$$

(Our above ϵ obtains by putting $m = n-1$ and $B_\mu = \cos\frac{\mu\pi}{n}$.)

We denote the mean of any function

$$f(x_1, \cdots, x_m)$$

over the above-mentioned spherical surface (the $x_1, \cdots, x_m$ being equidistributed over it) by

$$\overline{f(x_1, \quad \cdot, x_m)}$$

Our primary objective is to determine the moments of this distribution

$$M_p = \overline{\gamma^p} = \overline{\left(\sum_{\mu=1}^{m} B_\mu x_\mu^2\right)^p}, \qquad (p = 0, 1, 2, \cdots).$$

Let us write Σ_m for the (m − 1-dimensional) area of the above-mentioned spherical surface (of the unit sphere in m-dimensional Euclidean space).

Now we form the function

$$f(z) = \int_{-\infty}^{\infty} \cdot \quad \int_{-\infty}^{\infty} e^{z \sum_{\mu=1}^{m} B_\mu x_\mu^2} e^{-\sum_{\mu=1}^{m} x_\mu^2} \, dx_1 \cdots dx_m .$$

(This integral, as well as all others which we are going to derive from it, is obviously convergent, as long as z is sufficiently small More precisely this is true when

$$| z | \cdot \mathrm{Max} \, (| B_1 |, \cdot \quad , | B_m |) \leqq 1$$

We shall use them only in the neighborhood of $z = 0$.) Now clearly

$$\begin{aligned}
\left\{\frac{d^p}{dz^p} f(z)\right\}_{z=0} &= \left\{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\sum_{\mu=1}^{m} B_\mu x_\mu^2\right)^p e^{z \sum_{\mu=1}^{m} B_\mu x_\mu^2} e^{-\sum_{\mu=1}^{m} x_\mu^2} \, dx_1 \cdots dx_m\right\}_{z=0} \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\sum_{\mu=1}^{m} B_\mu x_\mu^2\right)^p e^{-\sum_{\mu=1}^{m} x_\mu^2} \, dx_1 \cdots dx_m \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} M_p \left(\sum_{\mu=1}^{m} x_\mu^2\right)^p e^{-\sum_{\mu=1}^{m} x_\mu^2} \, dx_1 \cdots dx_m \\
&= \int_0^{\infty} M_p r^{2p} e^{-r^2} \Sigma_m r^{m-1} \, dr \\
&= \Sigma_m M_p \int_0^{\infty} e^{-r^2} r^{2p+m-1} \, dr \quad [5] \\
&= \tfrac{1}{2} \Sigma_m M_p \int_0^{\infty} e^{-u} u^{p+\frac{1}{2}m-1} \, du \\
&= \tfrac{1}{2} \Sigma_m M_p \Gamma\left(p + \frac{m}{2}\right).
\end{aligned}$$

[5] Introduce the new integration variable $u = r^2$.

On the other hand

$$\begin{aligned} f(z) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\sum_{\mu=1}^{m}(1-B_\mu z)x_\mu^2}\, dx_1 \cdots dx_m \\ &= \prod_{\mu=1}^{m} \int_{-\infty}^{\infty} e^{-(1-B_\mu z)x_\mu^2}\, dx_\mu \quad [6] \\ &= \prod_{\mu=1}^{m} \tfrac{1}{2}(1 - B_\mu z)^{-\frac{1}{2}} \cdot 2\int_0^{\infty} e^{-u} u^{-\frac{1}{2}}\, du \\ &= \prod_{\mu=1}^{m} \tfrac{1}{2}(1 - B_\mu z)^{-\frac{1}{2}} \cdot 2\Gamma(\tfrac{1}{2}) \\ &= \Gamma(\tfrac{1}{2})^m \mathfrak{P}(z)^{-\frac{1}{2}}, \end{aligned}$$

where

$$\mathfrak{P}(z) = \prod_{\mu=1}^{m} (1 - B_\mu z).$$

Thus

$$\tfrac{1}{2}\Sigma_m M_p \Gamma\left(p + \frac{m}{2}\right) = \Gamma\left(\frac{1}{2}\right)^m \left\{\frac{d^p}{dz^p}\, \mathfrak{P}(z)^{-\frac{1}{2}}\right\}_{z=0}.$$

For $p = 0$ this becomes, since $M_0 = 1$, $\mathfrak{P}(0) = 1$,

$$\frac{1}{2}\, \Sigma_m \Gamma\left(\frac{m}{2}\right) = \Gamma\left(\frac{1}{2}\right)^m.$$

Dividing the former equation by the latter gives, since

$$\frac{\Gamma\left(p + \frac{m}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} = \frac{m}{2}\left(\frac{m}{2} + 1\right) \cdots \left(\frac{m}{2} + p - 1\right),$$

$$M_p = \frac{1}{\frac{m}{2}\left(\frac{m}{2} + 1\right) \cdots \left(\frac{m}{2} + p - 1\right)} \left\{\frac{d^p}{dz^p}\, \mathfrak{P}(z)^{-\frac{1}{2}}\right\}_{z=0}$$

In order to make a practical use of the above formula, we compute

$$\begin{aligned} \ln\left(\mathfrak{P}(z)^{-\frac{1}{2}}\right) &= -\tfrac{1}{2} \sum_{\mu=1}^{m} \ln(1 - B_\mu z) \\ &= -\tfrac{1}{2} \sum_{\mu=1}^{m} \sum_{l=1}^{\infty} -\frac{1}{l} B_\mu^l z^l \\ &= \sum_{l=1}^{\infty} \frac{1}{2l} \left(\sum_{\mu=1}^{m} B_\mu^l\right) z^l. \end{aligned}$$

[6] Introduce the new integration variable $u = (1 - B_\mu z)r^2$.

Write

$$\alpha_l = \frac{1}{2l}\sum_{\mu=1}^{m} B_\mu^l,$$

then

$$\mathfrak{P}(z)^{-\frac{1}{2}} = e^{\alpha_1 z+\alpha_2 z^2+\alpha_3 z^3+} = 1 + \beta_1 z + \beta_2 z^2 + \beta_3 z^3 + \cdots,$$

and so

$$M_p = \frac{1\cdot 2\cdots p}{\frac{m}{2}\left(\frac{m}{2}+1\right)\cdots\left(\frac{m}{2}+p-1\right)}\beta_p.$$

Clearly

$$\beta_1 = \alpha_1,$$
$$\beta_2 = \alpha_2 + \tfrac{1}{2}\alpha_1^2,$$
$$\beta_3 = \alpha_3 + \alpha_1\alpha_2 + \tfrac{1}{6}\alpha_1^3,$$
$$\beta_4 = \alpha_4 + \tfrac{1}{2}\alpha_2^2 + \alpha_1\alpha_3 + \tfrac{1}{2}\alpha_1^2\alpha_2 + \tfrac{1}{24}\alpha_1^4.$$

In our application (cf. above)

$$B_{m+1-\mu} = -B_\mu.$$

This has the consequence that

$$\alpha_1 = \alpha_3 = \alpha_5 = \cdots = 0.$$

Thus the z functions we compute contain only even powers of z and consequently

$$\beta_1 = \beta_3 = \beta_5 = \cdots = 0,$$
$$M_1 = M_3 = M_5 = \cdots = 0,$$

and

$$\beta_2 = \alpha_2,$$
$$\beta_4 = \alpha_4 + \tfrac{1}{2}\alpha_2^2,$$
$$\beta_6 = \alpha_6 + \alpha_2\alpha_4 + \tfrac{1}{6}\alpha_2^3,$$
$$\beta_8 = \alpha_8 + \tfrac{1}{2}\alpha_4^2 + \alpha_2\alpha_6 + \tfrac{1}{2}\alpha_2^2\alpha_4 + \tfrac{1}{24}\alpha_2^4.$$

As mentioned before, we actually have $m = n - 1$ and $B_\mu \equiv \cos \frac{\mu\pi}{n}$. Consequently

$$\begin{aligned}
\alpha_l &= \frac{1}{2l} \sum_{\mu=1}^{n-1} \left\{ \cos \frac{\mu\pi}{n} \right\}^l = \frac{1}{2l} \sum_{\mu=1}^{n-1} \{ \tfrac{1}{2}(e^{i\mu\pi/n} + e^{-i\mu\pi/n}) \}^l \\
&= \frac{1}{2^{l+1} l} \sum_{\mu=1}^{n-1} \sum_{k=0}^{l} \binom{l}{k} e^{i(2k-l)\mu\pi/n} \text{ [7]} \\
&= \frac{1}{2^{l+1} l} \sum_{k=0}^{l} \binom{l}{k} \sum_{\mu=1}^{n-1} e^{i2\pi\mu(k-\frac{1}{2}l)/n} \\
&= \frac{1}{2^{l+1} l} \sum_{k=0}^{l} \binom{l}{k} \left\{ \sum_{\mu=0}^{n-1} e^{i2\pi\mu(k-\frac{1}{2}l)/n} - 1 \right\}.
\end{aligned}$$

The inner sum has obviously these values

$$\begin{aligned}
\sum_{\mu=0}^{n-1} e^{i2\pi\mu(k-\frac{1}{2}l)/n} &= n \quad \text{if } k - \tfrac{1}{2}l \text{ is divisible by } n \\
&= 0 \quad \text{otherwise.}
\end{aligned}$$

Also

$$\sum_{k=0}^{l} \binom{l}{k} \cdot (-1) = -2^l.$$

Consequently

$$\alpha_l = \frac{n}{2^{l+1} l} \sum_{k}{}' \binom{l}{k} - \frac{1}{2l},$$

where $\sum_k{}'$ extends over those $k = 0, \cdots, l$, for which $k - \frac{1}{2}l$ is divisible by n.

Let us now determine the k occurring in the following sum (as above, $k - \frac{1}{2}l$ is divisible by n) $\sum_k{}'$. $k = \frac{1}{2}l$ is clearly one of them. All others are of the form $k = \frac{1}{2}l \pm hn$, $h = 1, 2, \cdots$. The term contributed is the same for $+$ and for $-$, since

$$\binom{l}{\frac{1}{2}l + hn} = \binom{l}{\frac{1}{2}l - hn}.$$

So we have

$$\alpha_l \begin{cases} = 0, & \text{for } l \text{ odd,} \\ = \dfrac{1}{2l} \left\{ \dfrac{n}{2^l} \left[\binom{l}{\frac{1}{2}l} + 2 \sum_{h=1,2,\cdots} \binom{l}{\frac{1}{2}l - hn} \right] - 1 \right\}, & \text{for } l \text{ even.} \end{cases}$$

[7] As pointed out above, we need to consider only the even l

The number of terms which the sum $\sum_{h=1,2,\cdots}$ contributes depends on the comparative sizes of l and n. The number is clearly

$$0 \text{ for } \quad \tfrac{1}{2}l < n,$$

$$1 \text{ for } \quad n \leqq \tfrac{1}{2}l < 2n,$$

$$2 \text{ for } 2n \leqq \tfrac{1}{2}l < 3n,$$

$$\cdots\cdots\cdots\cdots$$

Explicit formulae follow.[8]

$$\alpha_1 = \alpha_3 = \alpha_5 = \alpha_7 = \alpha_9 = \cdots = 0,$$

$$\alpha_2 = \frac{n-2}{8}, \qquad (0 \text{ for } n = 1),$$

$$\alpha_4 = \frac{3n-8}{64}, \qquad (0 \text{ for } n = 1, 2),$$

$$\alpha_6 = \frac{5n-16}{192}, \qquad \left(0 \text{ for } n = 1, 2; \frac{1}{384}, n = 3\right),$$

$$\alpha_8 = \frac{35n-128}{2048}, \qquad \left(0 \text{ for } n = 1, 2; \frac{1}{2048}, n = 3; \frac{1}{128}, n = 4\right)$$

$$\beta_1 = \beta_3 = \beta_5 = \beta_7 = \beta_9 = \cdots = 0,$$

$$\beta_2 = \frac{n-2}{8}, \qquad (0 \text{ for } n = 1),$$

$$\beta_4 = \frac{n^2+2n-12}{128}, \qquad (0 \text{ for } n = 1, 2),$$

$$\beta_6 = \frac{n^3+12n^2+8n-168}{3072}, \qquad \left(0 \text{ for } n = 1, 2; \frac{5}{1024}, n = 3\right),$$

$$\beta_8 = \frac{n^4+28n^3+212n^2-64n-3696}{98304}, \qquad \left(0 \text{ for } n = 1, 2; \frac{35}{32768}, n = 3, \frac{35}{2048}, n = 4\right).$$

$$M_1 = M_3 = M_5 = M_7 = M_9 = \cdots = 0,$$

$$M_2 = \frac{8}{(n-1)(n+1)} \cdot \beta_2 = \frac{n-2}{(n-1)(n+1)}, \qquad (0 \text{ for } n = 1),$$

---

[8] The author wishes to express his thanks to Miss B. I. Hart for her kind help in carrying out these computations.

$$M_4 = \frac{384}{(n-1)(n+1)(n+3)(n+5)} \cdot \beta_4 = \frac{3(n^2+2n-12)}{(n-1)(n+1)(n+3)(n+5)},$$

(0 for $n = 1, 2$),

$$M_6 = \frac{46080}{(n-1)(n+1)(n+3)(n+5)(n+7)(n+9)} \cdot \beta_6$$

$$= \frac{15(n^3+12n^2+8n-168)}{(n-1)(n+1)(n+3)(n+5)(n+7)(n+9)},$$

(0 for $n = 1, 2$; $\frac{5}{1024}$, $n = 3$).

$$M_8 = \frac{10321920}{(n-1)(n+1)(n+3)(n+5)(n+7)(n+9)(n+11)(n+13)} \cdot \beta_8$$

$$= \frac{105(n^4+28n^3+212n^2-64n-3696)}{(n-1)(n+1)(n+3)(n+5)(n+7)(n+9)(n+11)(n+13)},$$

(0 for $n = 1, 2$; $\frac{35}{32768}$, $n = 3$; $\frac{1127}{131072}$, $n = 4$).

We conclude this section by obtaining asymptotic formulae for the distribution of ϵ when $n \to \infty$.

In this case our formulae show that all α_l (l even) behave asymptotically like constant multiples of n. It also appears from our formulae for the β_l (l even), that

$$\beta_l = \frac{1}{(\frac{1}{2}l)!}\,\alpha_2^{\frac{1}{2}l} + \text{a polynomial in } \alpha_2, \alpha_4, \cdots, \alpha_{l-2} \text{ of total order} \leqq \tfrac{1}{2}l - 1.$$

Consequently $\frac{1}{(\frac{1}{2}l)!}\,\alpha_2^{\frac{1}{2}l}$ is the dominant term in this expression, and so we have asymptotically

$$\beta_l \sim \frac{1}{(\frac{1}{2}l)!}\,\alpha_2^{\frac{1}{2}l} \sim \frac{1}{(\frac{1}{2}l)!}\left(\frac{n}{8}\right)^{\frac{1}{2}l}.$$

From this

$$M_l \sim \frac{l!}{\left(\frac{n}{2}\right)^l}\,\beta_l \sim \frac{l!}{(\frac{1}{2}l)!}\left(\frac{1}{2n}\right)^{\frac{1}{2}l}.$$

Now the normal distribution

$$c_1 e^{-y^2/2\sigma_1^2}\,dy, \qquad \left(c_1 = \frac{1}{\sigma_1\sqrt{2\pi}}\right),$$

with the mean 0 and the standard deviation σ_1 has the moments

$$m_l = \int_{-\infty}^{\infty} y^l\, c_1 e^{-y^2/2\sigma_1^2}\,dy.$$

This is clearly 0 for l odd, while for l even[9]

$$m_l = \sigma_1^{l+1} c_1 \cdot 2^{\frac{1}{2}(l+1)} \int_0^\infty e^{-u} u^{\frac{1}{2}(l-1)}\, du$$

$$= 2^{\frac{1}{2}(l+1)} \sigma_1^{l+1} c_1 \Gamma\left(\frac{l+1}{2}\right).$$

For $l = 0$ this becomes, since $m_0 = 1$,

$$1 = 2^{\frac{1}{2}} \sigma_1 c_1 \Gamma(\tfrac{1}{2}).$$

Dividing the former equation by the latter gives, since

$$\frac{\Gamma\left(\frac{l+1}{2}\right)}{\Gamma(\frac{1}{2})} = \frac{1}{2}\cdot\frac{3}{2}\cdots\frac{l-1}{2},$$

$$m_l = 1\cdot 3\cdots(l-1)\sigma_1^l = \frac{l!}{2^{\frac{1}{2}l}(\frac{1}{2}l)!}\sigma_1^l = \frac{l!}{(\frac{1}{2}l)!}\left(\frac{\sigma_1^2}{2}\right)^{\frac{1}{2}l}.$$

Comparing the formulae for M_l and for m_l shows that $M_l \sim m_l$ if $\frac{1}{2n} = \frac{\sigma_1^2}{2}$, $\sigma_1 = \sqrt{\frac{1}{n}}$. So we see:

For $n \to \infty$ the distribution of ϵ becomes asymptotically normal, with the mean 0 and the standard deviation $\sigma_1 = \sqrt{\frac{1}{n}}$. (The same result could be obtained by applying the general theorems of Liapounoff and others.)

5. The distribution law, general discussion. We return to the quantity γ, defined at the beginning of the preceding section, of which our ϵ is a special case. We wish to obtain direct information concerning the distribution law of this γ.

Since a permutation of the B_μ is permissible, we arrange them such that

$$B_1 \geqq B_2 \geqq \cdots \geqq B_m.$$

(In the special case $\gamma = \epsilon$, the $B_\mu = \cos\frac{\mu\pi}{n}$ are given in this arrangement.)

The distribution of γ covers obviously the interval

$$B_1 \geqq y \geqq B_m.$$

And if not $B_1 = \cdots = B_m$, i.e. if $B_1 > B_m$, which we assume to be the case, then we have obviously a continuous distribution law for γ in this interval. We denote it by $\omega(y)\, dy$.

---

[9] Introduce the new integration variable $u = \frac{y^2}{2\sigma_1^2}$.

Assume for the moment that $B_m > 0$. Then the quantity

$$\gamma^{-\frac{1}{2}m} = \left(\sum_{\mu=1}^{m} B_\mu x_\mu^2\right)^{-\frac{1}{2}m},$$

is bounded, and we can therefore form its mean value. This is the $-\frac{m}{2}$ moment of γ (cf. the beginning of the preceding section)

$$M_{-\frac{1}{2}m} = \overline{\gamma^{-\frac{1}{2}m}} = \overline{\left(\sum_{\mu=1}^{m} B_\mu x_\mu^2\right)^{-\frac{1}{2}m}} = \int_{B_m}^{B_1} y^{-\frac{1}{2}m}\,\omega(y)\,dy.$$

With any two $a > b > 0$ $\left(\text{we shall have } \frac{a}{b} \to \infty \text{ subsequently}\right)$ form the quantity

$$\begin{aligned} \iota(a, b) &= \int\cdots\int_{a^2 \geqq \sum_{\mu=1}^{m} x_\mu^2 \geqq b^2} \left(\sum_{\mu=1}^{m} x_\mu^2\right)^{-\frac{1}{2}m} dx_1 \cdots dx_m \quad [10] \\ &= \int_b^a r^{-m} \cdot \Sigma_m\, r^{m-1}\, dr = \Sigma_m \int_b^a \frac{dr}{r} \\ &= \Sigma_m \ln \frac{a}{b}. \end{aligned}$$

Consider next

$$\begin{aligned} s(a, b) &= \int\cdots\int_{a^2 \geqq \sum_{\mu=1}^{m} \frac{1}{B_\mu} x_\mu^2 \geqq b^2} \left(\sum_{\mu=1}^{m} x_\mu^2\right)^{-\frac{1}{2}m} dx_1 \cdots dx_m \quad [11] \\ &= \int\cdots\int_{a^2 \geqq \sum_{\mu=1}^{m} x_\mu^2 \geqq b^2} \left(\sum_{\mu=1}^{m} B_\mu x_\mu^2\right)^{-\frac{1}{2}m} \sqrt{\prod_{\mu=1}^{m} B_\mu}\, dx_1 \cdots dx_m \\ &= \int\cdots\int_{a^2 \geqq \sum_{\mu=1}^{m} x_\mu^2 \geqq b^2} M_{-\frac{1}{2}m} \left(\sum_{\mu=1}^{m} x_\mu^2\right)^{-\frac{1}{2}m} \sqrt{\prod_{\mu=1}^{m} B_\mu}\, dx_1 \cdots dx_m \\ &= M_{-\frac{1}{2}m} \sqrt{\prod_{\mu=1}^{m} B_\mu}\, \iota(a, b). \end{aligned}$$

---

[10] Concerning this transformation to polar coordinates and the quantity Σ_m cf. the first part of the preceding section.

[11] Replace each variable x_μ by $\sqrt{B_\mu}\, x_\mu$.

On the other hand, a comparison of their respective integration domains makes it clear that

$$t(B_m a, B_1 b) \leqq s(a, b) \leqq t(B_1 a, B_m b).$$

Thus

$$\Sigma_m \ln \frac{B_m a}{B_1 b} \leqq M_{-\frac{1}{m}} \sqrt{\prod_{\mu=1}^{m} B_\mu} \cdot \Sigma_m \ln \frac{a}{b} \leqq \Sigma_m \ln \frac{B_1 a}{B_m b},$$

i.e.

$$\frac{1}{\sqrt{\prod_{\mu=1}^{m} B_\mu}} \frac{\ln \frac{a}{b} - \ln \frac{B_1}{B_m}}{\ln \frac{a}{b}} \leqq M_{-\frac{1}{m}} \leqq \frac{1}{\sqrt{\prod_{\mu=1}^{m} B_\mu}} \frac{\ln \frac{a}{b} + \ln \frac{B_1}{B_m}}{\ln \frac{a}{b}}.$$

Now let $\frac{a}{b} \to \infty$, then

$$M_{-\frac{1}{m}} = \frac{1}{\sqrt{\prod_{\mu=1}^{m} B_\mu}}$$

obtains, i.e

$$\overline{\gamma^{-\frac{1}{m}}} = \int_{B_m}^{B_1} y^{-\frac{1}{m}} \omega(y)\, dy = \frac{1}{\sqrt{\prod_{\mu=1}^{m} B_\mu}}.$$

We now drop the assumption $B_m > 0$. We consider instead a real number z with $z < B_m$ Replace each B_μ by $B_\mu - z$ Then the one with $\mu = m$ becomes > 0. And γ is obviously replaced by $\gamma - z$. Consequently our above equation is now valid in the form

$$\overline{(\gamma - z)^{-\frac{1}{m}}} = \int_{B_m}^{B_1} (y - z)^{-\frac{1}{m}} \omega(y)\, dy = \frac{1}{\sqrt{\prod_{\mu=1}^{m} (B_\mu - z)}}.$$

Let now z be a complex variable. The second term of the above equation is a (locally) analytical function of z, except in the (real) interval $B_1 \geqq z \geqq B_m$ The third term, too, is a (locally) analytical function of z, except at the (real) points $B_1, \cdots, B_m$. Consequently both are one-valued analytical functions of z in the simply connected domain which obtains from the complex z plane by exclusion of the (real) half line

$$z \geqq B_m.$$

Hence the equation

$$\int_{B_m}^{B_1} (y - z)^{-\frac{1}{m}} \omega(y)\, dy = \frac{1}{\sqrt{\prod_{\mu=1}^{m} (B_\mu - z)}}, \tag{1}$$

which holds for all (real) $z < B_m$, remains true for all complex z of the above domain.[12]

We observe next that $\omega(y)$ is an analytical function of y in $B_1 \geqq y \geqq B_m$, whenever $y \neq B_1, \cdots, B_m$. This is easily established by using any multiple integral expression for $\omega(y)$ which, while hard to evaluate explicitly, puts this analyticity into evidence.[13]

---

[12] $(y - z)^{-1/m}$ and the factors $(B_\mu - z)^{-1}$ of $\dfrac{1}{\sqrt{\prod_{\mu=1}^{m} (B_\mu - z)}}$ are those branches of these analytical functions which are (real and) > 0 when z is (real and) $< B_\mu$. When m is even (as it will be, cf. below) the domain of analyticity is somewhat more extended, but we need not discuss this.

[13] The computation which follows gives the desired analyticity in a simple way, and also makes it clear why the analyticity fails at $y = B_1, \cdots, B_m$.

Consider the $y \neq B_1, \cdots, B_m$ in $B_1 \geqq y \geqq B_m$. The probability of $\gamma \leqq y$ is $p(y) = \int_{B_m}^{y} \omega(y)\, dy$, and we may establish its analyticity instead of that of $p'(y) = \omega(y)$.

Obviously $p(y)$ is equally the probability of $\sum_{\mu=1}^{m} B_\mu x_\mu^2 \leqq y \sum_{\mu=1}^{m} x_\mu^2$, if the $x_1, \cdots, x_m$ are equidistributed over a spherical surface $\sum_{\mu=1}^{m} x_\mu^2 = r^2$, with any given $r > 0$.

Our hypotheses concerning y imply $B_\nu > y > B_{\nu+1}$ for a suitable $\nu = 1, \cdots, m-1$. Consider now the expression

$$f(y) = \int \cdots \int_{\sum_{\mu=1}^{m} B_\mu x_\mu^2 \leqq y \sum_{\mu=1}^{m} x_\mu^2} e^{-\sum_{\mu=1}^{m} x_\mu^2}\, dx_1 \cdots dx_m$$

Transforming to polar coordinates, we obtain

$$f(y) = \int_0^\infty e^{-r^2} \cdot \Sigma_m\, p(y) r^{m-1}\, dr = \Sigma_m \int_0^\infty e^{-r^2} r^{m-1}\, dr \cdot p(y).$$

(Σ_m as before.) Hence it suffices to establish the analyticity of $f(y)$. Now on the other hand

$$f(y) = \int \cdots \int_{\sum_{\mu=1}^{\nu} (B_\mu - y) x_\mu^2 \leqq \sum_{\mu=\nu+1}^{m} (y - B_\mu) x_\mu^2} e^{-\sum_{\mu=1}^{m} x_\mu^2}\, dx_1 \cdots dx_m$$

$$= \frac{1}{\sqrt{\prod_{\mu=1}^{m} |B_\mu - y|}} \int \cdots \int_{\sum_{\mu=1}^{\nu} w_\mu^2 \leqq \sum_{\mu=\nu+1}^{m} w_\mu^2} e^{-\sum_{\mu=1}^{m} w_\mu^2 / |B_\mu - y|}\, dw_1 \cdots dw_m .$$

(We introduced the new variables $w_\mu = \sqrt{|B_\mu - y|}\, x_\mu$.) And this expression is clearly analytical in y, since $B_\nu > y > B_{\nu+1}$.

We shall need only the fact that $\omega(y)$ possesses $\frac{1}{2}m$ continuous derivatives at these places. (m will be assumed to be even, cf. below.) Its behavior at $y = B_1, \cdots, B_m$ will follow from our subsequent results in all cases where we need it.

In order to determine $\omega(y)$ from (1), as we now propose to do, it is very convenient to assume that m is even. We therefore make this assumption, and shall maintain it throughout most of what follows.

Consider a $y_0 \neq B_1, \cdots, B_m$ in $B_1 \geqq y_0 \geqq B_m$. Then $B_v > y > B_{v+1}$ for a suitable $v = 1, \cdot\ \cdot, m-1$. Now put

$$z = y_0 + it \qquad (t \text{ real and } > 0),$$

form (1), take the imaginary parts of both sides, and let $t \to 0$.

Consider first the left-hand side of (1). Since $\omega(y)$ possesses $\frac{1}{2}m$ continuous derivatives at $y = y_0$, we have

$$\omega(y) = \sum_{k=0}^{\frac{1}{2}m-1} \theta_k (y - y_0)^k + e(y)(y - y_0)^{\frac{1}{2}m}$$

with a bounded $e(y)$. Clearly

$$\theta_k = \frac{1}{k!}\left\{\frac{d^k}{dy^k}\,\omega(y)\right\}_{y=y_0}$$

Thus, since $\omega(y)$ is real, all θ_k are real and $e(y)$ is also real

Compute now the contribution of each one of the $\frac{1}{2}m + 1$ terms in the above expression for $\omega(y)$ to the imaginary part of the left-hand side of (1).

The last term, $e(y)\cdot(y - y_0)^{\frac{1}{2}m}$, gives

$$\Im \int_{B_m}^{B_1} (y - y_0 - it)^{-\frac{1}{2}m} e(y)(y - y_0)^{\frac{1}{2}m}\, dy = \Im \int_{B_m}^{B_1} \left(\frac{y - y_0}{y - y_0 - it}\right)^{\frac{1}{2}m} e(y)\, dy.$$

The integrand is uniformly bounded, and so the reality conditions cause the entire expression to $\to 0$ for $t \to 0$ Hence the contribution of this term is zero for $t \to 0$.

The other $\dfrac{m}{2}$ terms correspond to $k = 0, 1, \cdots, \dfrac{m}{2} - 1$, the k term being

$$\Im \int_{B_m}^{B_1} (y - y_0 - it)^{-\frac{1}{2}m} \cdot \theta_k (y - y_0)^k \cdot dy$$

$$= \theta_k \Im \int_{B_m}^{B_1} \frac{(y - y_0)^k}{(y - y_0 - it)^{\frac{1}{2}m}}\, dy$$

$$= \theta_k \Im \int_{B_m}^{B_1} \frac{\sum_{h=0}^{k} \binom{k}{h} (it)^h (y - y_0 - it)^{k-h}}{(y - y_0 - it)^{\frac{1}{2}m}}\, dy$$

$$= \theta_k \sum_{h=0}^{k} \binom{k}{h} \Im \left\{(it)^h \int_{B_m}^{B_1} (y - y_0 - it)^{k-h-\frac{1}{2}m}\, dy\right\}$$

The exponent $k - h - \frac{m}{2}$ in the integral is always $\leqq \binom{m}{2} - 1 - 0 - \frac{m}{2} = -1$, and it is $= -1$ if and only if $k = \frac{m}{2} - 1$, $h = 0$. Consider first a term where this is not the case, i.e. where the exponent $k - h - \frac{m}{2} < -1$. For such a term the expression $\Im\{\cdots\}$ becomes

$$\Im(it)^h \cdot \frac{1}{k - h - \frac{m}{2} + 1} \{(y - y_0 - it)^{k-h-\frac{1}{2}m+1}\}_{y=B_m}^{y=B_1}.$$

For $t \to 0$ the last factors are bounded and real, and so the entire expression $\to 0$. for $h = 0$ because of the reality conditions, for $h > 0$ because of $(it)^h \to 0$. Thus only the term $k = \frac{m}{2} - 1$, $h = 0$ can contribute something else than zero for $t \to 0$.

Now this term is equal to

$$\theta_{\frac{1}{2}m-1}\Im\{\ln(y - y_0 - it)\}_{y=B_m}^{y=B_1},$$

and for $t \to 0$ this converges to

$$\theta_{\frac{1}{2}m-1}\Im(i\pi) = \pi\theta_{\frac{1}{2}m-1} = \frac{\pi}{\left(\frac{m}{2} - 1\right)!}\left\{\frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}}\omega(y)\right\}_{y=y_0}.$$ [14]

Thus the imaginary part of the entire left-hand side of (1) converges for $t \to 0$ to this expression.

The right-hand side of (1) is easier to discuss. The imaginary part under consideration is now

$$\Im \frac{1}{\sqrt{\prod_{\mu=1}^{m}(B_\mu - y_0 - it)}} = \Im \prod_{\mu=1}^{m}(B_\mu - y_0 - it)^{-\frac{1}{2}}.$$

Considering[12] (its y is our $y_0 + it$), this converges for $t \to 0$ to

$$\Im \prod_{\mu=1}^{\nu}(B_\mu - y_0)^{-\frac{1}{2}} \prod_{\mu=\nu+1}^{m} i(y_0 - B_\mu)^{-\frac{1}{2}} = \Im i^{m-\nu} \frac{1}{\sqrt{\prod_{\mu=1}^{m}|B_\mu - y_0|}}.$$ [15]

---

[14] This evaluation $\{\ln(y - y_0 - it)\}_{y=B_m}^{y=B_1} \to i\pi$ is based on $t > 0$, and the fact that y moves on the real axis from B_m to B_1. It has no connection with[12].

[15] The square roots of the (real and) > 0 quantities

$$B_\mu - y_0 \ (\mu = 1, \cdots, \nu), \quad y_0 - B_\mu \ (\mu = \nu + 1, \cdots, m), \quad \text{and} \quad \prod_{\mu=1}^{m}|B_\mu - y_0|$$

are taken to be > 0.

If v (hence $m - v$) is even, then this is zero. If v (hence $m - v$) is odd, then this is equal to $(-1)^{\frac{1}{2}(m-v-1)} \dfrac{1}{\sqrt{\prod_{\mu=1}^{m} |B_\mu - y_0|}}$. Thus (1) becomes the following equation.

$$\frac{\pi}{\left(\frac{m}{2}-1\right)!}\left\{\frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}}\omega(y)\right\}_{y=y_0} \begin{cases} = 0 & \text{if } v \text{ is even,} \\ = (-1)^{\frac{1}{2}(m-v-1)} \dfrac{1}{\sqrt{\prod_{\mu=1}^{m} |B_\mu - y_0|}} & \text{if } v \text{ is odd.} \end{cases}$$

Simplifying this, and writing y for y_0, and also restating the definition of v gives

$$\frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}}\omega(y) \begin{cases} = 0 & \text{if } v \text{ is even,} \\ = (-1)^{\frac{1}{2}(m-v-1)} \dfrac{\left(\frac{m}{2}-1\right)!}{\pi} \dfrac{1}{\sqrt{\prod_{\mu=1}^{m} |B_\mu - y|}} & \text{if } v \text{ is odd,} \end{cases} \tag{2}$$

$$B_v > y > B_{v+1}, \quad v = 1, \cdots, m-1.$$

Observe finally, that if we put

$$\mathfrak{A}(y) = \prod_{\mu=1}^{m} (y - B_\mu),$$

then this product has v factors < 0 ($\mu = 1, \cdots, v$), while the others are > 0. So

$$\mathfrak{A}(y) \gtrless 0 \quad \text{for} \quad v \begin{matrix} \text{even} \\ \text{odd} \end{matrix},$$

and in the latter case

$$\prod_{\mu=1}^{m} |B_\mu - y| = -\mathfrak{A}(y).$$

It is clear how we may now rewrite (2)

We are now in a position to determine the behavior of $\omega(y)$ at $y = B_1, \cdots, B_m$ too, since we know how its $\frac{m}{2} - 1$-th derivative behaves in the immediate vicinity of these places. (2) shows that it is singular there, and that the nature of the singularity depends on the number of the μ, for which B_μ is equal to the y in question, i e. on the multiplicity of this root of our polynomial $\mathfrak{A}(y)$.

In our actual application (to $\gamma = \epsilon$, cf. the beginning of this section) the

B_μ are pairwise different, i.e all root multiplicities of $\mathfrak{A}(y)$ are equal to one. A further special case, which has a certain interest of its own, is when the B_μ are equal two by two, but otherwise different, i.e. all root multiplicities of $\mathfrak{A}(y)$ are equal to two. In the discussion which follows we shall therefore assume that one or the other of these two cases occurs

In the first case $\frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}}\omega(y)$ has on each side of a $y = B_\mu$ one of these two behaviors: It is identically zero, or it is singular, of the type $\frac{1}{\sqrt{|B_\mu - y|}}$. Thus it is at any rate integrable. Consequently $\frac{d^{\frac{1}{2}m-2}}{dy^{\frac{1}{2}m-2}}\omega(y)$ is continuous on each side of $y = B_\mu$, i.e. for both $y = B_\mu \pm 0$. Successive integrations give now that all $\frac{d^k}{dy^k}\omega(y)$, $k = 0, 1, \cdots, \frac{m}{2} - 2$, are continuous for both $y = B_\mu \pm 0$.

In the second case we have $B_1 = B_2 > B_3 = B_4 > \cdots > B_{m-1} = B_m$. So the ν with $B_\nu > y > B_{\nu+1}$ is necessarily even, and $\frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}}\omega(y)$ is identically zero for all of (2). Consequently $\frac{d^{\frac{1}{2}m-2}}{dy^{\frac{1}{2}m-2}}\omega(y)$ is again continuous on each side of $y = B_\mu$, i.e. for both $y = B_\mu \pm 0$. Successive integrations show again that all $\frac{d^k}{dy^k}\omega(y)$, $k = 0, 1, \cdots, \frac{m}{2} - 2$, are continuous for both $y = B_\mu \pm 0$.

We must therefore discuss only how much the $\frac{d^k}{dy^k}\omega(y)$, $k = 0, 1, \cdots, \frac{m}{2} - 2$, change from $y = B_\mu - 0$ to $y = B_\mu + 0$.

Let us return to the procedure by which we derived (2) from (1). We put again

$$z = y_0 + it \qquad (t \text{ real and } > 0)$$

and let $t \to \infty$. But we consider now (1) itself (and not merely its imaginary part), and we choose a $y_0 = B_\nu$.

Consider first the left-hand side of (1), always disregarding terms which stay bounded for $t \to 0$. Then we can replace the integral $\int_{B_m}^{B_1}$ of (1) by any $\int_{B_\nu - a}^{B_\nu + a}$ with any fixed $a > 0$, and this is equal to

$$\int_{B_\nu - a}^{B_\nu - 0} + \int_{B_\nu + 0}^{B_\nu + a}.$$

We choose this $a > 0$ so small that no $B_\mu \neq B_\nu$ lies between $B_\nu - a$ and $B_\nu + a$. I.e. all $\frac{d^k}{dy^k}\omega(y)$, $k = 0, 1, \cdots, \frac{m}{2} - 2$, are continuous from $B_\nu - a$ to $B_\nu - 0$ and also from $B_\nu + 0$ to $B_\nu + a$.

This being the case, we can evaluate the above sum of two integrals by $\frac{m}{2} - 1$ successive partial integrations. Thus we get

$$-\left\{\sum_{k=0}^{\frac{1}{2}m-2} \frac{\left(\frac{m}{2}-2-k\right)!}{\left(\frac{m}{2}-1\right)!}(y - B_v - it)^{-\frac{1}{2}m+1+k} \frac{d^k}{dy^k}\omega(y)\right\}_{y=B_v-a}^{y=B_v-0}$$

$$-\left\{\sum_{k=0}^{\frac{1}{2}m-2} \frac{\left(\frac{m}{2}-2-k\right)!}{\left(\frac{m}{2}-1\right)!}(y - B_v - it)^{-\frac{1}{2}m+1+k} \frac{d^k}{dy^k}\omega(y)\right\}_{y=B_v+0}^{y=B_v+a}$$

$$+ \frac{1}{\left(\frac{m}{2}-1\right)!}\int_{B_v-a}^{B_v+a}(y - B_v - it)^{-1} \frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}}\omega(y)\,dy.$$

In the first two lines the $y = B_v \pm a$ terms are bounded for $t \to 0$, therefore only the $y = B_v \pm 0$ terms need be kept. Then the first two lines give

$$\sum_{k=0}^{\frac{1}{2}m-2} \frac{\left(\frac{m}{2}-2-k\right)!}{\left(\frac{m}{2}-1\right)!}(-it)^{-\frac{1}{2}m+1+k}\left\{\frac{d^k}{dy^k}\omega(y)\right\}_{y=B_v-0}^{y=B_v+0},$$

up to terms which stay bounded for $t \to 0$. Consider now the third line. We know that the $\frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}}\omega(y)$ in its integrand can be majorized by $\frac{c_2}{\sqrt{|y - B_v|}}$ (for a suitable constant c_2, cf. our discussion preceding the present one). Thus the integral in question is majorized by

$$\int_{B_v-a}^{B_v+a} |y - B_v - it|^{-1} c_2 |y - B_v|^{-\frac{1}{2}}\,dy,$$

hence *a fortiori* by

$$\int_{-\infty}^{\infty} |y - B_v - it|^{-1} c_2 |y - B_v|^{-\frac{1}{2}}\,dy \quad [16]$$

$$= c_2 t^{-\frac{1}{2}} \int_{-\infty}^{\infty} |u - i|^{-1} |u|^{-\frac{1}{2}}\,du$$

$$= c_2 t^{-\frac{1}{2}} \int_{-\infty}^{\infty} \frac{du}{\sqrt{(u^2+1)\cdot|u|}} \quad [17]$$

$$= c_2 \int_0^{\infty} \frac{dv}{\sqrt{v^4+1}}\, t^{-\frac{1}{2}}.$$

---

[16] Introduce the new integration variable $u = \frac{y - B_v}{t}$.

[17] Introduce the new integration variable $v = \sqrt{|u|}$

Since the last integration is obviously finite, the entire expression is $O(t^{-\frac{1}{2}})$ for $t \to 0$.

Consequently the left-hand side of (1) is equal to

$$\sum_{k=0}^{\frac{1}{2}m-2} \frac{\left(\frac{m}{2}-2-k\right)!}{\left(\frac{m}{2}-1\right)!} (-it)^{-\frac{1}{2}m+1+k} \left\{\frac{d^k}{dy^k}\omega(y)\right\}_{y=B_v-0}^{y=B_v+0} + O(t^{-1}),$$

for $t \to 0$. $\left(\text{For } B_v = B_1 \text{ or } B_m \text{ the } \frac{d^k}{dy^k}\omega(y) \text{ at } y = B_v + 0 \text{ or } B_v - 0, \text{ respectively, must obviously be taken to be zero.}\right)$

Consider now the right-hand side of (1).

We first suppose the B_μ are pairwise different. The right-hand side in question is $\frac{1}{\sqrt{\prod_{\mu=1}^{m}(B_\mu - B_v - it)}}$, i.e. $O(t^{-\frac{1}{2}})$.

Secondly let us consider $B_1 = B_2 > B_3 = B_4 > \cdots > B_{m-1} = B_m$. So we may assume $v = 2\lambda \left(\lambda = 1, \cdots, \frac{m}{2}\right)$. The right-hand side of (1) becomes now a rational function, $\frac{1}{\prod_{k=1}^{\frac{1}{2}m}(B_{2k} - z)}$. (The sign is determined by[12].) So in our case it is $\frac{1}{\prod_{k=1}^{\frac{1}{2}m}(B_{2k} - B_{2\lambda} - it)}$, i.e. $\frac{(-1)^{\frac{1}{2}m-\lambda}}{\prod_{k=1}^{\lambda-1}(B_{2k} - B_{2\lambda}) \prod_{k=\lambda+1}^{\frac{1}{2}m}(B_{2\lambda} - B_{2k})} \cdot (-it)^{-1} + O(1)$.

Comparing these with our above expression gives therefore (for $t \to 0$)

$$\sum_{k=0}^{\frac{1}{2}m-2} \frac{\left(\frac{m}{2}-2-k\right)!}{\left(\frac{m}{2}-1\right)!} (-it)^{-\frac{1}{2}m+1+k} \left\{\frac{d^k}{dy^k}\omega(y)\right\}_{y=B_v-0}^{y=B_v+0}$$

$$= O(t^{-\frac{1}{2}}) \text{ in the first case,}$$

$$= \frac{(-1)^{\frac{1}{2}m-\lambda}}{\prod_{k=1}^{\lambda-1}(B_{2k} - B_{2\lambda}) \cdot \prod_{k=\lambda+1}^{\frac{1}{2}m}(B_{2\lambda} - B_{2k})} (-it)^{-1} + O(t^{-\frac{1}{2}}) \text{ in the second case.}$$

In this formula the left-hand side is a polynomial in $(-it)^{-1}$. Hence the $O(t^{-\frac{1}{2}})$ terms on the right-hand side must vanish, and otherwise all powers of $-it$ must have the same coefficient on both sides. Consequently

$$\frac{\left(\frac{m}{2}-2-k\right)!}{\left(\frac{m}{2}-1\right)!} \left\{\frac{d^k}{dy^k}\omega(y)\right\}_{y=B_v-0}^{y=B_v+0}$$

must vanish, except in the second case for the one value of k with $-\frac{m}{2}+1+k=-1$, i.e. $k=\frac{m}{2}-2$. So, with this one exception, we have

$$\left\{\frac{d^k}{dy^k}\,\omega(y)\right\}_{y=B_\nu+0}=\left\{\frac{d^k}{dy^k}\,\omega(y)\right\}_{y=B_\nu-0}.$$

And in the exceptional case (second case, $\nu=2\lambda$)

$$\left\{\frac{d^{\frac{1}{2}m-2}}{dy^{\frac{1}{2}m-2}}\,\omega(y)\right\}_{y=B_\nu-0}^{y=B_\nu+0}=(-1)^{\frac{1}{2}m-\lambda}\left(\frac{m}{2}-1\right)!\,\frac{1}{\prod_{k=1}^{\lambda-1}(B_{2k}-B_{2\lambda})\prod_{k=\lambda+1}^{\frac{1}{2}m}(B_{2\lambda}-B_{2k})}.$$

Thus in the first case all derivatives $\frac{d^k}{dy^k}\,\omega(y)$, $k=0,1,\cdots,\frac{m}{2}-2$, are continuous even at $y=B_1,\cdots,B_m$.

In the second case the same is true for $k=0,1,\cdots,\frac{m}{2}-3$, but the derivative with $k=\frac{m}{2}-2$ behaves differently for $y=B_1,\cdots,B_m$. Indeed, for $y=B_{2\lambda-1}=B_{2\lambda}\left(\lambda=1,\cdots,\frac{m}{2}\right)$ this derivative is continuous for both $y=B_{2\lambda}\pm 0$, but it increases from $B_{2\lambda}-0$ to $B_{2\lambda}+0$ by

$$(-1)^{\frac{1}{2}m-\lambda}\left(\frac{m}{2}-1\right)!\,\frac{1}{\prod_{k=1}^{\lambda-1}(B_{2k}-B_{2\lambda})\prod_{k=\lambda+1}^{\frac{1}{2}m}(B_{2\lambda}-B_{2k})}.$$

(At $y=B_1+0$ and B_m-0 the $\frac{d^k}{dy^k}\,\omega(y)$ must be thought to continue with the value zero.)

These rules, together with (2), determine $\omega(y)$ completely.

6. First special case. We consider the first special case, where the B_μ are pairwise different. We immediately specialize further, to $\gamma=\epsilon$, i.e. $m=n-1$, $B_\mu=\cos\frac{\mu\pi}{n}$ $(\mu=1,\cdots,n-1)$. (Cf. the beginning of the preceding section.) Since m must be even, n must be odd. The rules of section 5 determine $\omega(y)$; in particular all derivatives $\frac{d^k}{dy^k}\,\omega(y)$, $k=0,1,\cdots,\frac{n-1}{2}-2$, are everywhere continuous, beginning and ending with zero at $y=B_1$ and B_{n-1}, respectively.

In the even intervals

$$B_2\geqq y\geqq B_3,\qquad B_4\geqq y\geqq B_5,\cdots,B_{n-3}\geqq y\geqq B_{n-2},$$

the derivative $\frac{d^{\frac{1}{2}(n-1)-1}}{dy^{\frac{1}{2}(n-1)-1}}\omega(y)$ is zero, i.e. $\omega(y)$ is a polynomial of degree $\frac{1}{2}(n-1)-2$. In the odd intervals

$$B_1 \geqq y \geqq B_2\,, \qquad B_3 \geqq y \geqq B_4\,, \cdots, B_{n-2} \geqq y \geqq B_{n-1}\,,$$

we have

$$\frac{d^{\frac{1}{2}(n-1)-1}}{dy^{\frac{1}{2}(n-1)-1}}\omega(y) = \pm\frac{(\frac{1}{2}[n-1]-1)!}{\pi}\frac{1}{\sqrt{-\mathfrak{A}(y)}}$$

(the sign $\pm$ is alternating $(-1)^{\frac{1}{2}(n-1)-1}, (-1)^{\frac{1}{2}(n-1)-2}, \cdots, +$), where

$$\mathfrak{A}(y) = \prod_{\mu=1}^{n-1}\left(y - \cos\frac{\mu\pi}{n}\right).$$

Another expression for $\mathfrak{A}(y)$ may be found by the following method.

Clearly $$\frac{\sin(n\varphi)}{\sin\varphi} = \frac{e^{in\varphi}-e^{-in\varphi}}{e^{i\varphi}-e^{-i\varphi}} = \sum_{\mu=0}^{n-1} e^{i(n-1-2\mu)\varphi}$$

is a polynomial of $\cos\varphi = \frac{1}{2}(e^{i\varphi}+e^{-i\varphi})$ of degree $n-1$, with the highest coefficient 2^{n-1}. For $\varphi = \frac{\mu\pi}{n}$, $\mu = 1, \cdots, n-1$, $\sin(n\varphi) = 0$, $\sin\varphi \neq 0$, hence $\frac{\sin(n\varphi)}{\sin\varphi}$, as a polynomial in $\cos\varphi$, has the same roots as $\mathfrak{A}(y)$. $\mathfrak{A}(y)$ is a polynomial of degree $n-1$ with the highest coefficient 1. Consequently

$$\mathfrak{A}(\cos\varphi) = \frac{1}{2^{n-1}}\frac{\sin(n\varphi)}{\sin\varphi}\,.$$

This formula allows one to compute $\mathfrak{A}(y)$ quickly, examples are

$$n = 3\colon\ \mathfrak{A}(y) = y^2 - \tfrac{1}{4},$$
$$n = 5\colon\ \mathfrak{A}(y) = y^4 - \tfrac{3}{4}y^2 + \tfrac{1}{16},$$
$$n = 7.\ \mathfrak{A}(y) = y^6 - \tfrac{5}{4}y^4 + \tfrac{3}{8}y^2 - \tfrac{1}{64}.$$

The number of odd intervals, on which integrations must be carried out, is $\frac{1}{2}(n-1)$, but since those which are symmetric with respect to 0 require the same computations, only $\frac{1}{4}(n-1)$ or $\frac{1}{4}(n+1)$ must be considered. So there are $1, 1, 2, \cdots$ such intervals for $n = 3, 5, 7, \cdots$ respectively. The integrals are first elementary (arcsin), then elliptic, then hyperelliptic.

Numerical computations for $n = 3$ are immediate; for $n = 5, 7$ they have been carried out with considerable precision by B. I. Hart.

At $y = B_\mu$, $\frac{d^{\frac{1}{2}(n-1)-1}}{dy^{\frac{1}{2}(n-1)-1}}\omega(y)$ has a singularity of the type $\frac{1}{\sqrt{|y-B_\mu|}}$ (cf. the end of section 5), while all $\frac{d^k}{dy^k}\omega(y)$, $k = 0, 1, \cdots, \frac{1}{2}(n-1)-2$, are continuous.

At $y = B_1$ and B_{n-1}, in particular, they are zero. Hence it follows by successive integrations that the order of vanishing of $\frac{d^k}{dy^k}\omega(y)$, $k = 0, 1, \cdots, \frac{1}{2}(n-1) - 2$ at $y = B_1$ and B_{n-1} is $(\frac{1}{2}(n-1) - 1) - k - \frac{1}{2} = \frac{n}{2} - 2 - k$. In particular for $k = 0$ we find that at its maximum and at its minimum $\left(B_1 \text{ and } B_{n-1}, \text{ i.e. } \pm\cos\frac{\pi}{n}\right)$ the order of vanishing of $\omega(y)$ is $\frac{n}{2} - 2$ [18]

Since $\omega(y)$ has this property, and since it is obviously an even function of y, R. H. Kent has suggested approximating it by a series expansion of the form

$$\omega(y) = \sum_{h=0}^{\infty} a_h \left(\cos^2\frac{\pi}{n} - y^2\right)^{\frac{1}{2}n-2+h} \tag{3}$$

Computations by B. I. Hart, not yet published, have shown that even the use of the first four terms ($h = 0, 1, 2, 3$, the a_h being determined by the condition of normalization and by the first three even moments of the actual distribution given in section 4) give excellent approximations. The use of the formula (3) suggests itself likewise for even values of n.

7. Second special case. We consider now the second special case, where $B_1 = B_2 > B_3 = B_4 > \cdots > B_{m-1} = B_m$ This has no immediate bearing on our original problem (cf. the preceding section), but we shall nevertheless discuss it for the two following reasons First, it is hoped that the reader will find an independent interest in the simple and complete results which can be obtained in this case. Second, there are various modifications of our original problem, which lead to this case. For example let the $x_1, \cdots, x_n$ in our original problem, as described in section 1, be complex numbers instead of real ones, replacing all squares by absolute value squares. Then one verifies easily that all characteristic values $\lambda_1, \cdots, \lambda_{n-1}$ are doubled, and so our first case goes over into our second case (This amounts to replacing our quadratic forms by Hermitian forms, cf [4]) It is easy to imagine two-dimensional problems where this set-up is natural.

We put $C_\lambda = B_{2\lambda-1} = B_{2\lambda}$ for $\lambda = 1, \cdots, \frac{m}{2}$, so that $C_1 > C_2 > \cdots > C_{\frac{1}{2}m}$ are the only restrictions imposed

Every y in $B_1 \geqq y \geqq B_m$, i.e. in $C_1 \geqq y \geqq C_{\frac{1}{2}m}$, lies in an interval $C_\lambda \geqq y \geqq C_{\lambda+1}$ i.e $B_{2\lambda} \geqq y \geqq B_{2\lambda+1}$. That is the v of (2) is always even, and so $\frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}}\omega(y)$ is zero in every one of these intervals. Therefore $\omega(y)$ is a polynomial of degree $\frac{m}{2} - 2$ in every one of these intervals. We have already shown that $\omega(y)$ is

[18] We omit the simple discussion of $n = 3$, which must be excluded from this result

not the same polynomial in each interval. Thus $\omega(y)$ is represented by $\frac{m}{2} - 1$ polynomials of degree $\frac{m}{2} - 2$ in the $\frac{m}{2} - 1$ intervals

$$C_1 \geq y \geq C_2, \quad C_2 \geq y \geq C_3, \quad \cdots, \quad C_{\frac{m}{2}-1} \geq y \geq C_{\frac{m}{2}}.$$

We could try to obtain explicit expressions for these polynomials by a direct application of the results at the close of section 5. A characterization of the distribution can, however, be obtained in a more elegant way by an indirect procedure.

Consider an arbitrary function $\mathfrak{F}(y)$. We wish to express its mean

$$\overline{\mathfrak{F}}(y) = \int_{C_{\frac{m}{2}}}^{C_1} \mathfrak{F}(y)\omega(y)\, dy.$$

If we can do this for all $\mathfrak{F}(y)$ then the distribution is completely characterized.

We select first an $\frac{m}{2} - 1$-fold primitive function of $\mathfrak{F}(y)$, i.e. a function $\mathfrak{G}(y)$ with

$$\frac{d^{\frac{m}{2}-1}}{dy^{\frac{m}{2}-1}}\mathfrak{G}(y) = \mathfrak{F}(y).$$

Of course $\mathfrak{G}(y)$ is determined only up to an additive polynomial of degree $\frac{m}{2} - 2$ in y.

Now the above expectation value becomes

$$\begin{aligned}\overline{\mathfrak{F}}(y) &= \int_{C_{\frac{m}{2}}}^{C_1} \frac{d^{\frac{m}{2}-1}}{dy^{\frac{m}{2}-1}}\mathfrak{G}(y)\omega(y)\, dy \\ &= \sum_{\lambda=1}^{\frac{m}{2}-1} \int_{C_{\lambda+1}+0}^{C_\lambda - 0} \frac{d^{\frac{m}{2}-1}}{dy^{\frac{m}{2}-1}}\mathfrak{G}(y)\omega(y)\, dy.\end{aligned}$$

Since all $\frac{d^k}{dy^k}\omega(y)$, $k = 0, 1, \cdots, \frac{m}{2} - 2$, are continuous from $C_{\lambda+1} + 0$ to $C_\lambda - 0$ for all $\lambda = 1, \cdots, \frac{m}{2} - 1$, we can evaluate each integral of the above sum by $\frac{m}{2} - 1$ successive partial integrations. Thus the following expression obtains:

$$\sum_{\lambda=1}^{\frac{m}{2}-1}\left\{\sum_{k=0}^{\frac{m}{2}-2}(-1)^k \frac{d^{\frac{m}{2}-k-2}}{dy^{\frac{m}{2}-k-2}}\mathfrak{G}(y)\,\frac{d^k}{dy^k}\omega(y)\right\}_{y=C_{\lambda+1}+0}^{y=C_\lambda - 0} + (-1)^{\frac{m}{2}-1}\int_{C_{\frac{m}{2}}}^{C_1}\mathfrak{G}(y)\,\frac{d^{\frac{m}{2}-1}}{dy^{\frac{m}{2}-1}}\omega(y)\, dy.$$

Considering the definition of $\mathfrak{G}(y)$ as an $\frac{m}{2} - 1$-fold primitive function, the $\frac{d^{k'}}{dy^{k'}}\mathfrak{G}(y)$, $k' = 0, 1, \cdots, \frac{m}{2} - 2$, are everywhere continuous. This corresponds

to $k' = \frac{m}{2} - k - 2, k = 0, 1, \cdots, \frac{m}{2} - 2$ Hence the first line can be rewritten as

$$-\sum_{\lambda=1}^{\frac{1}{2}m} \sum_{k=0}^{\frac{1}{2}m-2} (-1)^k \left\{ \frac{d^{\frac{1}{2}m-k-2}}{dy^{\frac{1}{2}m-k-2}} \mathfrak{G}(y) \right\}_{y=C_\lambda} \left\{ \frac{d^k}{dy^k} \omega(y) \right\}_{y=C_\lambda-0}^{y=C_\lambda+0}.$$

(For $C_\lambda = C_1$ or $C_{\frac{1}{2}m}$ the $\frac{d^k}{dy^k} \omega(y)$ at $y = C_1 + 0$ or $C_{\frac{1}{2}m} - 0$, respectively, must obviously be taken to be zero.) Owing to the results of section 5 all terms with $k = 0, 1, \cdots, \frac{m}{2} - 3$ vanish, and the term with $k = \frac{m}{2} - 2$ gives

$$-\sum_{\lambda=1}^{\frac{1}{2}m} (-1)^{\frac{1}{2}m-2} \mathfrak{G}(C_\lambda)(-1)^{\frac{1}{2}m-\lambda} \left(\frac{m}{2} - 1\right)! \frac{1}{\prod_{k=1}^{\lambda-1} (C_k - C_\lambda) \prod_{k=\lambda+1}^{\frac{1}{2}m} (C_\lambda - C_k)}$$

$$= \sum_{\lambda=1}^{\frac{1}{2}m} (-1)^{\lambda-1} \left(\frac{m}{2} - 1\right)! \frac{1}{\prod_{k=1}^{\lambda-1} (C_k - C_\lambda) \prod_{k=\lambda+1}^{\frac{1}{2}m} (C_\lambda - C_k)} \mathfrak{G}(C_\lambda).$$

The second line vanishes, since $\frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}} \omega(y)$ is zero everywhere, as observed above.

Finally

$$\overline{\mathfrak{F}(y)} = \sum_{\lambda=1}^{\frac{1}{2}m} (-1)^{\lambda-1} \left(\frac{m}{2} - 1\right)! \frac{1}{\prod_{k=1}^{\lambda-1} (C_k - C_\lambda) \prod_{k=\lambda+1}^{\frac{1}{2}m} (C_\lambda - C_k)} \mathfrak{G}(C_\lambda).$$

For

$$\mathfrak{B}(z) = \prod_{k=1}^{\frac{1}{2}m} (z - C_k)$$

we have

$$\left\{ \frac{d}{dz} \mathfrak{B}(z) \right\}_{z=C_\lambda} = \prod_{k=1}^{\frac{1}{2}m}{}_{(k \neq \lambda)} (C_\lambda - C_k)$$

$$= (-1)^{\lambda-1} \prod_{k=1}^{\lambda-1} (C_k - C_\lambda) \prod_{k=\lambda+1}^{\frac{1}{2}m} (C_\lambda - C_k).$$

Therefore the above formula can also be written

$$\overline{\mathfrak{F}(y)} = \left(\frac{m}{2} - 1\right)! \sum_{\lambda=1}^{\frac{1}{2}m} \frac{\mathfrak{G}(C_\lambda)}{\left\{ \frac{d}{dz} \mathfrak{B}(z) \right\}_{z=C_\lambda}}.$$

Observe that the right-hand side of the above formula (which can also be easily expressed in terms of determinants) is a well-known approximate ex-

pression for $\frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}}\,\mathfrak{G}(y)$, as a (repeated) difference quotient of the values $\mathfrak{G}(C_\lambda)$, $\lambda = 1, \cdots, \frac{m}{2}$. It is therefore very satisfactory that this expression gives the mean of

$$\mathfrak{F}(y) = \frac{d^{\frac{1}{2}m-1}}{dy^{\frac{1}{2}m-1}}\,\mathfrak{G}(y).$$

Appendix. We return to the normal distribution of $x_1, \cdots, x_n$ as described in section 1, and to the quantities $\overline{s^2}$, $\overline{\delta^2}$, η given there. We denote means with respect to that distribution by $\overline{(\cdots)}$.

It was observed by B. I. Hart and mentioned by J. D. Williams[3] by comparing the known expressions for their moments, that every moment of $\eta = \frac{\delta^2}{s^2}$ is the quotient of the corresponding moments of δ^2 and of s^2. That is

$$\overline{\left(\frac{\delta^{2p}}{s^{2p}}\right)} = \frac{\overline{\delta^{2p}}}{\overline{s^{2p}}}, \qquad (p = 0, 1, 2, \cdots).$$

This indicates some kind of independence relation involving δ^2 and s^2. The considerations which follow are intended to clarify this situation.

The above relation may be written

$$\overline{s^{2p}}\;\overline{\eta^p} = \overline{s^{2p}\eta^p},$$

or, more generally,

$$\overline{s^q}\;\overline{\eta^p} = \overline{s^q\eta^p}.$$

We shall prove this by showing that s and η are statistically independent.

We can, as in section 2, make the mean $\xi = 0$, i.e. obtain the $x_1, \cdots, x_n$ distribution law

$$c^n e^{-\sum_{\mu=1}^{n} x_\mu^2/2\sigma^2}\,dx_1 \cdots dx_n.$$

And then, again as in section 2, perform a linear orthogonal transformation, carrying $x_1, \cdots, x_n$ into, say $x_1', \cdots, x_n'$ which leaves the distribution law in its original form

$$c^n e^{-\sum_{\mu=1}^{n} x_\mu'^2/2\sigma^2}\,dx_1' \cdots dx_n',$$

and makes

$$s^2 = \frac{1}{n}\sum_{\mu=1}^{n-1} x_\mu'^2,$$

$$\eta = \frac{n}{n-1}\,\frac{\sum_{\mu=1}^{n-1} A_\mu x_\mu'^2}{\sum_{\mu=1}^{n-1} x_\mu'^2}.$$

Since x_n' does not occur in s^2, η we must use only the $x_1', \cdots, x_{n-1}'$ distribution law

$$c^{n-1} e^{-\sum_{\mu=1}^{n-1} x_\mu'^2/2\sigma^2} dx_1' \cdots dx_{n-1}'.$$

Now we introduce polar coordinates with respect to $x_1', \cdot\ \cdot, x_{n-1}'$. These consist of a radius r with

$$r^2 = \sum_{\mu=1}^{n-1} x_\mu'^2,$$

and $n - 2$ angular variables $\varphi_1, \cdots, \varphi_{n-2}$, which can be chosen in various ways, and which we need not describe more closely At any rate

$$dx_1' \cdots dx_{n-1}' = r^{n-2}\, dr w(\varphi_1, \cdots, \varphi_{n-2})\, d\varphi_1 \cdots d\varphi_{n-2}$$

where we need not determine the weight function $w(\varphi_1, \cdots, \varphi_{n-2})$. Consequently the distribution law is

$$c^{n-1} e^{-r^2/2\sigma^2} r^{n-2}\, dr w(\varphi_1, \cdots, \varphi_{n-2})\, d\varphi_1 \cdots d\varphi_{n-2}.$$

Thus the coordinate r and the coordinates $\varphi_1, \cdots, \varphi_{n-2}$ are independent of each other.

Next

$$s^2 = \frac{1}{n} r^2,$$

and η is a homogeneous function of $x_1', \cdots, x_{n-1}'$ of degree zero, i.e. it is independent of r. So s is a function of r alone, and η is a function of $\varphi_1, \cdots, \varphi_{n-2}$ alone. Consequently s and η likewise are independent.

Added in proof:

After this manuscript was completed, Dr. T. Koopmans informed the author of several results of his own, which he obtained in connection with other statistical investigations. They have many points of contact with this investigation, and will appear in the near future in the *Annals of Mathematical Statistics*. The author wishes to express his thanks to Dr. T. Koopmans for his communications.

SOME EXAMPLES OF ASYMPTOTICALLY MOST POWERFUL TESTS

By Abraham Wald[1]

Columbia University

1. Introduction. In a previous paper[2] the author gave the definition of an asymptotically most powerful test and has shown that the commonly used tests, based on the maximum likelihood estimate, are asymptotically most powerful.

In this paper some further examples of asymptotically most powerful tests will be given. Let us first restate the definition of an asymptotically most powerful test. Let $f(x, \theta)$ be the probability density of a variate x involving an unknown parameter θ. For testing the hypothesis $\theta = \theta_0$ by means of n independent observations $x_1, \cdots, x_n$ on x we have to choose a region of rejection W_n in the n-dimensional sample space. Denote by $P(W_n \mid \theta)$ the probability that the sample point $E = (x_1, \cdots, x_n)$ will fall in W_n under the assumption that θ is the true value of the parameter. For any region U_n of the n-dimensional sample space denote by $g(U_n)$ the greatest lower bound of $P(U_n \mid \theta)$. For any pair of regions U_n and T_n denote by $L(U_n, T_n)$ the least upper bound of

$$P(U_n \mid \theta) - P(T_n \mid \theta).$$

In all that follows we shall denote a region of the n-dimensional sample space by a capital letter with the subscript n.

Definition 1. A sequence $\{W_n\}$ $(n = 1, 2, \cdots, \text{ad inf.})$ of regions is said to be an asymptotically most powerful test of the hypothesis $\theta = \theta_0$ on the level of significance α if $P(W_n \mid \theta_0) = \alpha$ and if for any sequence $\{Z_n\}$ of regions for which $P(Z_n \mid \theta_0) = \alpha$ the inequality

$$\limsup_{n \to \infty} L(Z_n, W_n) \leq 0$$

holds.

Definition 2: A sequence $\{W_n\}$ $(n = 1, 2, \cdots, \text{ad inf.})$ of regions is said to be an asymptotically most powerful unbiased test of the hypothesis $\theta = \theta_0$ on the level of significance α if $P(W_n \mid \theta_0) = \lim_{n \to \infty} g(W_n) = \alpha$, and if for any sequence $\{Z_n\}$ of regions for which $P(Z_n \mid \theta_0) = \lim_{n \to \infty} g(Z_n) = \alpha$, the inequality

$$\limsup_{n \to \infty} L(Z_n, W_n) \leq 0$$

holds.

---

[1] Research under a grant-in-aid of the Carnegie Corporation of New York.

[2] "Asymptotically most powerful tests of statistical hypotheses," *Annals of Math. Stat.* Vol. 12 (1941).

Consider the expression

$$y_n(\theta) = \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{n} \frac{\partial}{\partial\theta} \log f(x_\alpha, \theta) \tag{1}$$

Let W_n' be the region defined by the inequality $y_n(\theta_0) \geq c_n'$, W_n'' defined by the inequality $y_n(\theta_0) \leq c_n''$, and W_n defined by the inequality $| y_n(\theta_0) | \geq c_n$, where the constants c_n', c_n'' and c_n are chosen such that

$$P(W_n' \mid \theta_0) = P(W_n'' \mid \theta_0) = P(W_n \mid \theta_0) = \alpha.$$

It will be shown in this paper that under certain restrictions on the probability density $f(x, \theta)$ the sequence $\{W_n'\}$ is an asymptotically most powerful test of the hypothesis $\theta = \theta_0$ if θ takes only values $\geq \theta_0$. Similarly $\{W_n''\}$ is an asymptotically most powerful test if θ takes only values $\leq \theta_0$. Finally $\{W_n\}$ is an asymptotically most powerful unbiased test if θ can take any real value.

Another example of an asymptotically most powerful unbiased test of the hypothesis $\theta = \theta_0$, as it will be shown, is the critical region of type A in the Neyman-Pearson theory of testing hypotheses. This fact gives a strong justification for the use of the critical region of type A.

2. Assumptions on the density function. Let ω be a subset of the real axis. Denote by θ^* a real variable which takes only values in ω and let θ be a variable which can take any real value. For any function $\psi(x)$ we denote by $E_\theta\psi(x)$ the expected value of $\psi(x)$ under the assumption that θ is the true value of the parameter, i.e.

$$E_\theta \psi(x) = \int_{-\infty}^{+\infty} \psi(x) f(x, \theta)\, dx.$$

For any x, for any positive δ and for any real value θ_1 denote by $\varphi_1(x, \theta_1, \delta)$ the greatest lower bound, and by $\varphi_2(x, \theta_1, \delta)$ the least upper bound of $\frac{\partial^2}{\partial\theta^2} \log f(x, \theta)$ in the interval $\theta_1 - \delta \leq \theta \leq \theta_1 + \delta$. In all that follows the symbol θ_i^*, for any integer i, will denote a value of θ^*, i.e., θ_i^* is a point of ω.

We say that a value θ lies in the ϵ-neighborhood of ω if there exists a value θ^* such that $| \theta - \theta^* | \leq \epsilon$.

Throughout the paper the following assumptions on $f(x, \theta)$ will be made.

ASSUMPTION 1. *For any pair of sequences* $\{\theta_n\}$ *and* $\{\theta_n^*\}$ $(n = 1, 2, \ldots, \text{ad inf.})$ *for which*

$$\lim_{n=\infty} E_{\theta_n} \frac{\partial}{\partial\theta} \log f(x, \theta_n^*) = 0$$

also

$$\lim_{n=\infty} (\theta_n - \theta_n^*) = 0.$$

Furthermore there exists a positive ϵ such that $E_\theta\left[\frac{\partial}{\partial\theta}\log f(x,\theta_1)\right]^2$ *is a bounded function of* θ *and* θ_1, $E_\theta\frac{\partial}{\partial\theta}\log f(x,\theta_1)$ *is a continuous function of* θ *and* θ_1 *and* $E_{\theta_1}\left[\frac{\partial}{\partial\theta}\log f(x,\theta_1)\right]^2 = d(\theta_1)$ *has a positive lower bound, where* θ_1 *can take any value in the* ϵ*-neighborhood of* ω.

ASSUMPTION 2: *There exists a positive* k_0 *such that* $E_{\theta_2}\varphi_1(x,\theta_1,\delta)$ *and* $E_{\theta_2}\varphi_2(x,\theta_1,\delta)$ *are uniformly continuous functions in the domain* D *defined as follows: the variables* θ_1 *and* θ_2 *may take any value in the* k_0*-neighborhood of* ω *and* δ *may take any value for which* $|\delta| \leq k_0$. *Furthermore it is assumed that*

$$E_{\theta_1}[\varphi_i(x,\theta_1,\delta)]^2, \qquad (i = 1, 2)$$

are bounded functions of θ_1, θ_2 *and* δ *in* D.

ASSUMPTION 3: *There exists a positive* k_0 *such that*

$$\int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta}f(x,\theta)\,dx = \int_{-\infty}^{+\infty}\frac{\partial^2}{\partial\theta^2}f(x,\theta)\,dx = 0$$

for all θ *in the* k_0*-neighborhood of* ω.

Assumption 3 means simply that we may differentiate with respect to θ under the integral sign. In fact,

$$\int_{-\infty}^{+\infty}f(x,\theta)\,dx = 1,$$

identically in θ. Hence

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}f(x,\theta)\,dx = \frac{\partial^2}{\partial\theta^2}\int_{-\infty}^{+\infty}f(x,\theta)\,dx = 0.$$

Differentiating under the integral sign we obtain the relations in Assumption 3.

ASSUMPTION 4: *There exists a positive* k_0 *and a positive* η *such that*

$$E_\theta\left[\frac{\partial}{\partial\theta}\log f(x,\theta)\right]^{2+\eta}$$

is a bounded function of θ *in the* k_0*-neighborhood of* ω.

3. Some propositions. PROPOSITION 1: *To any positive* β *there exists a positive* γ *such that*

$$\lim_{n\to\infty}P\left\{\frac{1}{\sqrt{n}}\,|\,y_n(\theta^*)\,| > \gamma \,|\, \theta\right\} = 1$$

uniformly in θ^* *and for all* θ *for which* $|\theta - \theta^*| \geq \beta$.

PROOF: From Assumption 1 it follows that $\left|E_\theta\frac{\partial}{\partial\theta}\log f(x,\theta^*)\right|$ has a positive

lower bound in the domain $|\theta - \theta^*| \geq \beta$. Since according to Assumption 1 $E_\theta\left[\frac{\partial}{\partial\theta}\log f(x, \theta^*)\right]^2$ is a bounded function of θ and θ^*, Proposition 1 easily follows.

PROPOSITION 2: *There exists a positive* ϵ *such that*

$$\lim_{n=\infty} P[y_n(\theta) < t \mid \theta] = N(t \mid \theta)$$

uniformly in t *and for all* θ *in the* ϵ*-neighborhood of* ω *where*

$$d(\theta) = -E_\theta \frac{\partial^2}{\partial\theta^2}\log f(x, \theta) = E_\theta\left[\frac{\partial}{\partial\theta}\log f(x, \theta)\right]^2 \tag{2}$$

and

$$N(t \mid \theta) = \frac{1}{\sqrt{2\pi d(\theta)}}\int_{-\infty}^{t} e^{-\frac{1}{2}v^2/d(\theta)}\,dv. \tag{3}$$

Proposition 2 follows easily from Assumptions 3 and 4 and the general limit theorems.

PROPOSITION 3: *There exists a positive* ϵ *such that for any bounded sequence* $\{\mu_n\}$

$$\lim_{n=\infty}\left\{P\left[y_n(\theta) < t \mid \theta + \frac{\mu_n}{\sqrt{n}}\right] - \int_{-\infty}^{t} e^{\mu_n v - \frac{1}{2}\mu_n^2 d(\theta)}\,dN(v \mid \theta)\right\} = 0$$

uniformly in t *and for all* θ *in the* ϵ*-neighborhood of* ω.

PROOF: We have

$$y_n\left(\theta + \frac{\mu_n}{\sqrt{n}}\right) = y_n(\theta) + \frac{\mu_n}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_\alpha \frac{\partial^2}{\partial\theta^2}\log f(x_\alpha, \theta_n') \tag{4}$$

where θ_n' lies in the interval $\left[\theta, \theta + \frac{\mu_n}{\sqrt{n}}\right]$. From Assumption 2 and the above equation we easily obtain

$$\lim_{n=\infty}\left\{P\left[y_n\left(\theta + \frac{\mu_n}{\sqrt{n}}\right) < t \mid \theta + \frac{\mu_n}{\sqrt{n}}\right] - P\left[y_n(\theta) - \mu_n d(\theta) < t \mid \theta + \frac{\mu_n}{\sqrt{n}}\right]\right\} = 0 \tag{5}$$

uniformly in t and for all θ in the ϵ-neighborhood of ω. From Proposition 2 and (5) we get

$$\lim_{n=\infty}\left\{\int_{-\infty}^{t} dN(v \mid \theta) - P\left[y_n(\theta) < t + \mu_n d(\theta) \,\middle|\, \theta + \frac{\mu_n}{\sqrt{n}}\right]\right\} = 0$$

or

$$\lim_{n=\infty}\left\{\int_{-\infty}^{t-\mu_n d(\theta)} dN(v \mid \theta) - P\left[y_n(\theta) < t \mid \theta + \frac{\mu_n}{\sqrt{n}}\right]\right\} = 0 \tag{6}$$

uniformly in t and for all θ in the ϵ-neighbourhood of ω. This proves Proposition 3

PROPOSITION 4: *There exists a positive ϵ such that for any positive γ and for any sequence $\{\mu_n\}$ for which* $\lim_{n=\infty} |\mu_n| = \infty$

$$\lim_{n=\infty} P\left\{|y_n(\theta^*)| > \gamma \mid \theta^* + \frac{\mu_n}{\sqrt{n}}\right\} = 1$$

uniformly in θ^*.

PROOF: If there exists a positive β such that $\left|\frac{\mu_n}{\sqrt{n}}\right| > \beta$ for almost all n, Proposition 4 follows from Proposition 1. Hence we have to consider only the case $\lim_{n=\infty} \frac{\mu_n}{\sqrt{n}} = 0$. Since

$$E_{\theta^*+(\mu_n/\sqrt{n})}\, y_n\left(\theta^* + \frac{\mu_n}{\sqrt{n}}\right) = 0,$$

we get from (4)

$$E_{\theta^*+(\mu_n/\sqrt{n})}[y_n(\theta^*)] + \mu_n E_{\theta^*+(\mu_n/\sqrt{n})} \frac{\sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta_n')}{n} = 0. \tag{7}$$

Since $\lim \frac{\mu_n}{\sqrt{n}} = 0$, we have on account of Assumption 2

$$\lim_{n=\infty} E_{\theta^*+(\mu_n/\sqrt{n})} \frac{\sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta_n')}{n} = E_{\theta^*} \frac{\partial^2}{\partial\theta^2} \log f(x, \theta^*)$$

$$= -E_{\theta^*}\left[\frac{\partial}{\partial\theta} \log f(x, \theta^*)\right]^2 = -d(\theta^*)$$

uniformly in θ^*. According to Assumption 1 $d(\theta^*)$ has a positive lower bound; hence on account of $\lim |\mu_n| = \infty$ we obtain from (7)

$$\lim_{n=\infty} |E_{\theta^*+(\mu_n/\sqrt{n})}\, y_n(\theta^*)| = \infty \tag{8}$$

uniformly in θ^*. The variance of $y_n(\theta^*)$ is equal to the variance of $\frac{\partial}{\partial\theta} \log f(x, \theta^*)$. On account of Assumption 1 the variance of $\frac{\partial}{\partial\theta} \log f(x, \theta^*)$ (under the assumption that $\theta^* + \frac{\mu_n}{\sqrt{n}}$ is the true value of the parameter) is a bounded function. Hence Proposition 4 is proved on account of (8).

PROPOSITION 5. *Let $\{W_n(\theta^*)\}$ be a sequence of regions of size α, i.e. $P[W_n(\theta^*) \mid \theta^*] = \alpha$, and let $V_n(\theta^*, y)$ be the region defined by the inequality*

$y_n(\theta^*) < y$. *Let* $U_n(\theta^*, y)$ *be the intersection of* $V_n(\theta^*, y)$ *and* $W_n(\theta^*)$ *and denote* $P[U_n(\theta^*, y) \mid \theta^*]$ *by* $F_n(y \mid \theta^*)$. *Denote furthermore* $P\left[W_n(\theta^*) \mid \theta^* + \frac{\mu}{\sqrt{n}}\right]$ *by* $G(\theta^*, \mu, n)$ *If* $\{\theta_n^*\}$ *and* $\{\mu_n\}$ *are two sequences such that* $\lim_{n=\infty} d(\theta_n^*) = d$; $\lim_{n=\infty} F_n(y \mid \theta_n^*) = F(y)$ *and* $\lim \mu_n = \mu$ *then*

$$\lim_{n=\infty} G(\theta_n^*, \mu_n, n) = \int_{-\infty}^{+\infty} e^{\mu y - \frac{1}{2}\mu^2 d} \, dF(y).$$

PROOF: Let $\lim \mu_n = \mu$ and consider the Taylor expansion

$$\sum_\alpha \log f\left(x_\alpha, \theta^* + \frac{\mu_n}{\sqrt{n}}\right) = \sum_\alpha \log f(x_\alpha, \theta^*) + \frac{\mu_n}{\sqrt{n}} \sum_\alpha \frac{\partial}{\partial\theta} \log f(x_\alpha, \theta^*) + \frac{1}{2}\frac{\mu_n^2}{n} \frac{\partial^2}{\partial\theta^2} \sum_\alpha \log f(x_\alpha, \theta_n') \tag{9}$$

where θ_n' lies in the interval $\left[\theta^*, \theta^* + \frac{\mu_n}{\sqrt{n}}\right]$. From this we easily get on account of Assumption 2 and the fact that $\{\mu_n\}$ is bounded

$$\log \prod_{\alpha=1}^{n} \frac{f\left(x_\alpha, \theta^* + \frac{\mu_n}{\sqrt{n}}\right)}{f(x_\alpha, \theta^*)} = \mu_n y_n(\theta^*) - \tfrac{1}{2}\mu_n^2 d(\theta^*) + \epsilon(\theta^*, n) \tag{10}$$

where for arbitrary positive η

$$\lim_{n=\infty} P\left\{|\epsilon(\theta^*, n)| < \eta \mid \theta^* + \frac{\mu_n}{\sqrt{n}}\right\} = 1 \tag{11}$$

uniformly in θ^*. Denote by $R_n(\theta^*)$ the region defined by

$$|\epsilon(\theta^*, n)| < \eta > 0. \tag{12}$$

On account of (11) we have

$$\lim_{n=\infty} P\left[R_n(\theta^*) \mid \theta^* + \frac{\mu_n}{\sqrt{n}}\right] = 1, \tag{13}$$

uniformly in θ^*. Denote the intersection of $R_n(\theta^*)$ and $W_n(\theta^*)$ by $Q_n(\theta^*)$, and the intersection of $R_n(\theta^*)$ and $U_n(\theta^*, y)$ by $T_n(\theta^*, y)$. Furthermore denote $P[T_n(\theta^*, y) \mid \theta^*]$ by $\bar{F}_n(y \mid \theta^*)$. Then we have

$$e^{-\eta} \int_{-\infty}^{t} e^{\mu_n y - \frac{1}{2}\mu_n^2 d(\theta^*)} \, d\bar{F}_n(y \mid \theta^*) \leq P\left[T_n(\theta^*, t) \mid \theta^* + \frac{\mu_n}{\sqrt{n}}\right] \leq e^{\eta} \int_{-\infty}^{t} e^{\mu_n y - \frac{1}{2}\mu_n^2 d(\theta^*)} \, d\bar{F}_n(y \mid \theta^*) \tag{14}$$

for all values of t and θ^*. Furthermore we obviously have

$$\lim_{n=\infty}\left\{G(\theta^*, \mu_n, n) - P\left[Q_n(\theta^*) \mid \theta^* + \frac{\mu_n}{\sqrt{n}}\right]\right\} = 0 \tag{15}$$

uniformly in θ^*, and

$$\lim_{n=\infty}[\bar{F}_n(t \mid \theta^*) - F_n(t \mid \theta^*)] = 0 \tag{16}$$

uniformly in θ^* and t. Since η may be chosen arbitrarily small, it follows from (14) and (15) that to any $\epsilon > 0$, η may be chosen such that

$$\limsup_{n=\infty}\left|G(\theta_n^*, \mu_n, n) - \int_{-\infty}^{+\infty} e^{\mu_n t - \frac{1}{2}\mu_n^2 d(\theta_n^*)}\, d\bar{F}_n(t \mid \theta_n^*)\right| \leqq \frac{\epsilon}{2} \tag{17}$$

for any sequence $\{\theta_n^*\}$

To each ϵ let L_ϵ be a positive number such that L_ϵ depends only on ϵ and

$$\int_{-\infty}^{-L_\epsilon} e^{\mu_n t - \frac{1}{2}\mu_n^2 d(\theta^*)}\, dN(t \mid \theta^*) + \int_{L_\epsilon}^{\infty} e^{\mu_n t - \frac{1}{2}\mu_n^2 d(\theta^*)}\, dN(t \mid \theta^*) \leq \frac{\epsilon}{2} \tag{18}$$

for all n and for all values of θ^*. Since $d(\theta^*)$ has a positive lower and a finite upper bound, it is easy to verify that such a L_ϵ exists. From (18) and Proposition 3 it follows

$$\limsup_{n\to\infty}\left\{P\left[y_n(\theta_n^*) < -L_\epsilon \mid \theta_n^* + \frac{\mu_n}{\sqrt{n}}\right] + P\left[y_n(\theta_n^*) > L_\epsilon \mid \theta_n^* + \frac{\mu_n}{\sqrt{n}}\right]\right\} \leq \frac{\epsilon}{2} \tag{19}$$

for any arbitrary sequence $\{\theta_n^*\}$. Since the difference $U_n(\theta^*, t_2) - U_n(\theta^*, t_1)$ is a subset of the difference $V_n(\theta^*, t_2) - V_n(\theta^*, t_1)$ and since $T_n(\theta^*, t_2) - T_n(\theta^*, t_1)$ is a subset of $U_n(\theta^*, t_2) - U_n(\theta^*, t_1)$ for $t_2 > t_1$, we get from (18) and (19)

$$\limsup_{n\to\infty}\left\{P\left[U_n(\theta_n^*, -L_\epsilon) \mid \theta_n^* + \frac{\mu_n}{\sqrt{n}}\right] + P\left[W_n(\theta_n^*) \mid \theta_n^* + \frac{\mu_n}{\sqrt{n}}\right] - P\left[U_n(\theta_n^*, L_\epsilon) \mid \theta_n^* + \frac{\mu_n}{\sqrt{n}}\right]\right\} \leq \frac{\epsilon}{2} \tag{20}$$

and

$$\limsup_{n\to\infty}\left\{P\left[T_n(\theta_n^*, -L_\epsilon) \mid \theta_n^* + \frac{\mu_n}{\sqrt{n}}\right] + P\left[Q_n(\theta_n^*) \mid \theta_n^* + \frac{\mu_n}{\sqrt{n}}\right] - P\left[T_n(\theta_n^*, L_\epsilon) \mid \theta_n^* + \frac{\mu_n}{\sqrt{n}}\right]\right\} \leq \frac{\epsilon}{2} \tag{21}$$

for any sequence $\{\theta_n^*\}$. On account of (14) we get from (21)

$$e^{-\eta}\limsup_{n\to\infty}\left\{\int_{-\infty}^{-L_\epsilon} e^{\mu_n t - \frac{1}{2}\mu_n^2 d(\theta_n^*)}\, d\bar{F}_n(t \mid \theta_n^*) + \int_{L_\epsilon}^{\infty} e^{\mu_n t - \frac{1}{2}\mu_n^2 d(\theta_n^*)}\, d\bar{F}_n(t \mid \theta_n^*)\right\} \leq \frac{\epsilon}{2}. \tag{22}$$

From (17) and (22) we obtain

$$\limsup_{n\to\infty} \left| G(\theta_n^*, \mu_n, n) - \int_{-L_\epsilon}^{L_\epsilon} e^{\mu_n t - \frac{1}{2}\mu_n^2 d(\theta_n^*)}\, d\bar{F}_n(t \mid \theta_n^*) \right| \leq \epsilon\left(\frac{1+e^\eta}{2}\right) \tag{23}$$

for any sequence $\{\theta_n^*\}$. Consider now the sequence $\{\theta_n^*\}$ which satisfies the conditions of Proposition 5. Since $F_n(t \mid \theta_n^*)$ converges to $F(t)$ uniformly in t, on account of (16) also $\bar{F}_n(t \mid \theta_n^*)$ converges to $F(t)$ uniformly in t Hence we obtain from (23)

$$\limsup_{n\to\infty} \left| G(\theta_n^*, \mu_n, n) - \int_{-L_\epsilon}^{L_\epsilon} e^{\mu t - \frac{1}{2}\mu^2 d}\, dF(t) \right| \leq \epsilon\left(\frac{1+e^\eta}{2}\right). \tag{24}$$

Since ϵ and η may be chosen arbitrarily small, Proposition 5 follows from (24).

4. Some theorems and corollaries. THEOREM 1. *Denote by $S_n(\theta^*)$ the region defined by the inequality $y_n(\theta^*) \geq A_n(\theta^*)$ where $A_n(\theta^*)$ is chosen such that $P[S_n(\theta^*) \mid \theta^*] = \alpha$. For any region $W_n(\theta^*)$ denote by $L_n[W_n(\theta^*)]$ the least upper bound of $P[W_n(\theta^*) \mid \theta] - P[S_n(\theta^*) \mid \theta]$ with respect to θ^* and θ, where θ is restricted to values $\geq \theta^*$ Then for any sequence $\{W_n(\theta^*)\}$ for which $P[W_n(\theta^*) \mid \theta^*] = \alpha$,*

$$\limsup_{n\to\infty} L_n[W_n(\theta^*)] \leq 0.$$

PROOF. Assume that Theorem 1 is not true Then there exists a sequence of integers $\{n'\}$, a sequence $\{\theta_{n'}^*\}$ and a sequence $\{\theta_{n'}\}$ $(\theta_{n'} \geq \theta_{n'}^*)$ such that

$$\lim_{n'\to\infty} \{P[W_{n'}(\theta_{n'}^*) \mid \theta_{n'}] - P[S_{n'}(\theta_{n'}^*) \mid \theta_{n'}]\} = \delta > 0 \tag{25}$$

On account of Proposition 2 and Assumption 2 the sequence $\{A_{n'}(\theta_{n'}^*)\}$ is bounded. Then it follows easily from (25) and Proposition 4 (taking in account that $E_\theta \frac{\partial}{\partial\theta} \log f(x, \theta^*) > 0$ for $\theta > \theta^*$

$$(\theta_{n'} - \theta_{n'}^*)\sqrt{n'} = \mu_{n'} > 0 \tag{26}$$

must be bounded. Denote by $\{n''\}$ a subsequence of $\{n'\}$ such that

$$\lim d(\theta_{n''}^*) = d \tag{27}$$

$$\lim \mu_{n''} = \mu, \quad \text{and} \tag{28}$$

$$\lim F_{n''}(t \mid \theta_{n''}^*) = F(t) \tag{29}$$

uniformly in t where

$$F_n(t \mid \theta^*) = P[U_n(\theta^*, t) \mid \theta^*]$$

and $U_n(\theta^*, t)$ is the intersection of $W_n(\theta^*)$ and the region $y_n(\theta^*) < t$ The existence of a subsequence $\{n''\}$ such that (29) holds follows from the fact that

$$F_n(t_2 \mid \theta^*) - F_n(t_1 \mid \theta^*) \leq \Phi_n(t_2 \mid \theta^*) - \Phi_n(t_1 \mid \theta^*) \quad \text{for} \quad t_2 > t_1, \tag{30}$$

and

$$\lim_{n=\infty} \Phi_n(t \mid \theta_n^{*\prime\prime}) = \frac{1}{\sqrt{2\pi d}} \int_{-\infty}^{t} e^{-\frac{1}{2}v^2/d}\, dv = N(t), \tag{31}$$

where $\Phi_n(t \mid \theta^*)$ denotes the probability $P[y_n(\theta^*) < t \mid \theta^*]$. Furthermore it can easily be shown that

$$\int_{-\infty}^{+\infty} dF(t) = \alpha. \tag{32}$$

On account of Proposition 5 we get from (25), (27), (28), (29), (30) and (31)

$$\int_{-\infty}^{+\infty} e^{\mu t - \frac{1}{2}\mu^2 d}\, dF(t) - \int_{A}^{\infty} e^{\mu t - \frac{1}{2}\mu^2 d}\, dN(t) = \delta, \tag{33}$$

where A denotes a value such that

$$\int_{A}^{\infty} dN(t) = \alpha.$$

It has been shown in a previous paper[3] that (33) leads to a contradiction. Hence Theorem 1 is proved.

THEOREM 2: *Denote by $S_n(\theta^*)$ the region defined by the inequality $y_n(\theta^*) \leq A_n(\theta^*)$ where $A_n(\theta^*)$ is chosen such that $P[S_n(\theta^*) \mid \theta^*] = \alpha$. For any region $W_n(\theta^*)$ denote by $L_n[W_n(\theta^*)]$ the least upper bound of*

$$P[W_n(\theta^*) \mid \theta] - P[S_n(\theta^*) \mid \theta]$$

with respect to θ^* and θ, where θ is restricted to values $\leq \theta^*$. Then for any sequence $\{W_n(\theta^*)\}$ for which $P[W_n(\theta^*) \mid \theta^*) = \alpha$,

$$\limsup_{n\to\infty} L_n[W_n(\theta^*)] \leq 0.$$

The proof is omitted, since it is analogous to that of Theorem 1.

THEOREM 3: *Let $\{W_n(\theta^*)\}$ be for each θ^* a sequence of regions for which $P[W_n(\theta^*) \mid \theta^*] = \alpha$ and $\lim_{n=\infty} g[W_n(\theta^*)] = \alpha$ uniformly in θ^*. Denote by $L_n[W_n(\theta^*)]$ the least upper bound of*

$$P[W_n(\theta^*) \mid \theta] - P[\mid y_n(\theta^*) \mid \geq A_n(\theta^*) \mid \theta]$$

with respect to θ and θ^*, where $A_n(\theta^*)$ is chosen such that

$$P[\mid y_n(\theta^*) \mid \geq A_n(\theta^*) \mid \theta^*] = \alpha.$$

Then

$$\limsup_{n\to\infty} L_n[W_n(\theta^*)] \leq 0.$$

---

[3] See p. 12 of the paper cited in [2].

PROOF. Denote $P[y_n(\theta^*) < t \mid \theta^*]$ by $\Phi_n(t \mid \theta^*)$ and denote by $F_n(t \mid \theta^*)$ the probability (under the hypothesis $\theta = \theta^*$) of the intersection of $W_n(\theta^*)$ with the region $y_n(\theta^*) < t$. Assume that Theorem 3 is not true. Then there exists a subsequence $\{n''\}$, a sequence $\{\theta^{*}_{n''}\}$ and a sequence $\{\theta_{n''}\}$ such that

$$\lim_{n''=\infty} d(\theta^{*}_{n''}) = d; \qquad \lim (\theta_{n''} - \theta^{*}_{n''})\sqrt{n''} = \lim \mu_{n''} = \mu;$$

$$\lim F_{n''}(t \mid \theta^{*}_{n''}) = F(t)$$

uniformly in t, and

$$\int_{-\infty}^{+\infty} e^{\mu t - \frac{1}{2}\mu^2 d}\, dF(t) - \int_{-\infty}^{-A} e^{\mu t - \frac{1}{2}\mu^2 d}\, dN(t) - \int_{A}^{\infty} e^{\mu t - \frac{1}{2}\mu^2 d}\, dN(t) = \delta \tag{34}$$

where A is a positive number such that

$$\int_{-\infty}^{-A} dN(t) = \frac{\alpha}{2}, \qquad \text{and} \qquad N(t) = \frac{1}{\sqrt{2\pi d}} \int_{-\infty}^{t} e^{-\frac{1}{2}v^2/d}\, dv.$$

This can be proved in the same way as (33) has been proved. The author has shown in a previous paper[4] that (34) leads to a contradiction. Hence Theorem 3 is proved.

THEOREM 4. *Denote by $A_n(\theta^*)$ the region of type[5] A of size α for testing the hypothesis $\theta = \theta^*$. Denote by $B_n(\theta^*)$ the region $\mid y_n(\theta^*) \mid \geq C_n(\theta^*)$ where $C_n(\theta^*)$ is determined such that*

$$P[\mid y_n(\theta^*) \mid \geq C_n(\theta^*) \mid \theta^*] = \alpha.$$

Then, under the assumption that $E_\theta\left[\frac{\partial^2}{\partial\theta^2} \log f(x, \theta^*)\right]^2$ is bounded,

$$\lim_{n=\infty} \{P[A_n(\theta^*) \mid \theta] - P[B_n(\theta^*) \mid \theta]\} = 0$$

uniformly in θ and θ^*.

PROOF. The region $A_n(\theta^*)$ is given by the inequality[6]

$$\left[\sum_\alpha \frac{\partial}{\partial\theta} \log f(x_\alpha, \theta^*)\right]^2 + \sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta^*) \geq k'_n(\theta^*)\left[\sum_\alpha \frac{\partial}{\partial\theta} \log f(x_\alpha, \theta^*)\right] + k''_n(\theta^*), \tag{35}$$

where $k'_n(\theta^*)$ and $k''_n(\theta^*)$ are chosen such that $A_n(\theta^*)$ should be unbiased and of size α. The inequality (35) can be written also in the form

$$[y_n(\theta^*)]^2 + \frac{1}{n}\sum_\alpha \frac{\partial^2}{\partial\theta^2} \log f(x_\alpha, \theta^*) \geq l'_n(\theta^*) y_n(\theta^*) + l''_n(\theta^*). \tag{36}$$

---

[4] See p. 14 of the paper cited in [2].

[5] Neyman, J. and Pearson, E. S., "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Mem.*, Vol. 1.

[6] See the paper cited in [5].

Let $\{\mu_n\}$ be a bounded sequence. From Assumption 2 it follows that for any positive ϵ

$$P\left\{\left|\frac{1}{n}\sum_{\alpha}\frac{\partial^2}{\partial\theta^2}\log f(x_\alpha,\theta^*)+d(\theta^*)\right|<\epsilon\mid\theta^*+\frac{\mu_n}{\sqrt{n}}\right\}=1 \tag{37}$$

uniformly in θ^*. Since (37) holds for arbitrarily small ϵ, we get easily on account of Proposition 3

$$\lim_{n=\infty}\left\{P\left[A_n(\theta^*)\mid\theta^*+\frac{\mu_n}{\sqrt{n}}\right]-P\left[A_n'(\theta^*)\mid\theta^*+\frac{\mu_n}{\sqrt{n}}\right]\right\}=0 \tag{38}$$

uniformly in θ^*, where $A_n'(\theta^*)$ is defined by

$$[y_n(\theta^*)]^2\geq l_n'(\theta^*)y_n(\theta^*)+l_n''(\theta^*)+d(\theta^*). \tag{39}$$

Since $A_n(\theta^*)$ is unbiased and of size α, we have on account of (38) and (39)

$$\lim l_n'(\theta^*)=0 \quad \text{and} \tag{40}$$

$$\lim l_n''(\theta^*)+d(\theta^*)=\lambda(\theta^*)>0 \tag{41}$$

uniformly in θ^*, where $\lambda(\theta^*)$ is given by the condition

$$\frac{1}{\sqrt{2\pi d(\theta^*)}}\int_{-\sqrt{\lambda(\theta^*)}}^{+\sqrt{\lambda(\theta^*)}}e^{-\frac{1}{2}t^2/d(\theta^*)}\,dt=\alpha. \tag{42}$$

Inequality (39) is obviously equivalent to the simultaneous inequalities:

$$y_n(\theta^*)\leq c_n'(\theta^*) \qquad \text{and} \qquad y_n(\theta^*)\geq c_n''(\theta^*)$$

where $c_n'(\theta^*)$ and $c_n''(\theta^*)$ are the roots of the equation in $y_n(\theta^*)$

$$[y_n(\theta^*)]^2=l_n'(\theta^*)y_n(\theta^*)+l_n''(\theta^*)+d(\theta^*).$$

Since

$$\lim c_n'(\theta^*)=-\sqrt{\lambda(\theta^*)} \qquad \text{and} \qquad \lim c_n''(\theta^*)=+\sqrt{\lambda(\theta^*)}$$

uniformly in θ^*, from Proposition 3 it follows that

$$\lim_{n=\infty}\left\{P\left[A_n(\theta^*)\mid\theta^*+\frac{\mu_n}{\sqrt{n}}\right]-\int_{-\infty}^{-\sqrt{\lambda(\theta^*)}}e^{\mu_n t-\frac{1}{2}\mu_n^2 d(\theta^*)}\,dN(t\mid\theta^*)-\int_{+\sqrt{\lambda(\theta^*)}}^{\infty}e^{\mu_n t-\frac{1}{2}\mu_n^2 d(\theta^*)}\,dN(t\mid\theta^*)\right\}=0 \tag{43}$$

uniformly in θ^*.

Now let us consider a sequence $\{\nu_n\}$ such that $\lim|\nu_n|=\infty$ and $\lim\frac{\nu_n}{\sqrt{n}}=0$.

We shall prove that

$$P\left[A_n(\theta^*) \mid \theta^* + \frac{\nu_n}{\sqrt{n}}\right] = 1 \tag{44}$$

uniformly in θ^*. Since $E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(x, \theta^*)\right]^2$ is assumed to be bounded,

$$E_{\theta^*+(\nu_n/\sqrt{n})}\left[\frac{\partial^2}{\partial\theta^2}\log f(x, \theta^*)\right] \tag{45}$$

and

$$E_{\theta^*+(\nu_n/\sqrt{n})}\left[\frac{\partial^2}{\partial\theta^2}\log f(x, \theta^*)\right]^2 \tag{46}$$

are bounded functions of θ^* and n. We get by Taylor expansion

$$\sum_\alpha \frac{\partial}{\partial\theta}\log f(x_\alpha, \theta^*) = \sum_\alpha \frac{\partial}{\partial\theta}\log f\left(x_\alpha, \theta^* + \frac{\nu_n}{\sqrt{n}}\right) - \frac{\nu_n}{\sqrt{n}}\sum_\alpha \frac{\partial^2}{\partial\theta^2}\log f(x_\alpha, \bar{\theta}_n^*) \tag{47}$$

where $\bar{\theta}_n^*$ lies in $\left[\theta^*, \theta^* + \frac{\nu_n}{\sqrt{n}}\right]$. Hence

$$E_{\theta^*+(\nu_n/\sqrt{n})}[y_n(\theta^*)] = -\nu_n E_{\theta^*+(\nu_n/\sqrt{n})}\left[\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\sum_\alpha \log f(x_\alpha, \bar{\theta}_n^*)\right]. \tag{48}$$

From Assumption 2 and $\lim |\nu_n| = \infty$ it follows that the absolute value of the right hand side of (48) converges to ∞. Hence

$$\lim |E_{\theta^*+\nu_n/\sqrt{n}}[y_n(\theta^*)]| = \infty.$$

Since on account of Assumption 1

$$E_{\theta^*+(\nu_n/\sqrt{n})}\left[\frac{\partial}{\partial\theta}\log f(x_\alpha, \theta^*)\right]^2$$

is a bounded function of n and θ^*, also the variance of $y_n(\theta^*)$ (under the assumption that $\theta = \theta^* + \nu_n/\sqrt{n}$ is the true value of the parameter) is a bounded function of n and θ^* Hence for any arbitrary large constant C

$$\lim P\left[|y_n(\theta^*)| \geq C \mid \theta^* + \frac{\nu_n}{\sqrt{n}}\right] = 1, \tag{49}$$

uniformly in θ^*. The equation (44) follows easily from (36), (40), (41), (45), (46) and (49).

Consider a sequence $\{\rho_n\}$ such that $\left|\frac{\rho_n}{\sqrt{n}}\right| > \beta > 0$ for all n. Then it follows easily from Proposition 1 that for any arbitrary C

$$\lim P\left[\,|\,y_n(\theta^*)\,| \geq C \,|\, \theta^* + \frac{\rho_n}{\sqrt{n}}\right] = 1 \tag{50}$$

uniformly in θ^*. Since $E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(x_\alpha, \theta^*)\right]^2$ is assumed to be bounded, and therefore also $E_\theta \frac{\partial^2}{\partial\theta^2}\log f(x, \theta^*)$ is bounded, there exists a finite g such that

$$\lim P\left\{\left|\frac{1}{n}\sum_\alpha \frac{\partial^2}{\partial\theta^2}\log f(x_\alpha, \theta^*)\right| < g \,|\, \theta^* + \frac{\rho_n}{\sqrt{n}}\right\} = 1 \tag{51}$$

uniformly in θ^*. From (36), (40), (41), (50) and (51) it follows

$$\lim P\left[A_n(\theta^*) \,|\, \theta^* + \frac{\rho_n}{\sqrt{n}}\right] = 1 \tag{52}$$

uniformly in θ^*. Since on account of Propositions 3 and 4, the relations (43), (44) and (52) hold if we substitute $B_n(\theta^*)$ for $A_n(\theta^*)$, Theorem 4 is proved.

If Assumptions 1–4 are fulfilled for the set ω consisting of the single point $\theta = \theta_0$, then we get from Theorems 1–4 the following corollaries:

COROLLARY 1. *Let W_n' be the region defined by the inequality $y_n(\theta_0) \geq c_n'$; W_n'' defined by the inequality $y_n(\theta_0) \leq c_n''$, and W_n defined by the inequality $|\,y_n(\theta_0)\,| \geq c_n$, where the constants c_n', c_n'' and c_n are chosen such that*

$$P(W_n' \,|\, \theta_0) = P(W_n'' \,|\, \theta_0) = P(W_n \,|\, \theta_0) = \alpha.$$

Then $\{W_n'\}$ is an asymptotically most powerful test of the hypothesis $\theta = \theta_0$ if θ takes only values $\geq \theta_0$. Similarly $\{W_n''\}$ is an asymptotically most powerful test if θ takes only values $\leq \theta_0$. Finally $\{W_n\}$ is an asymptotically most powerful unbiased test if θ can take any real value.

COROLLARY 2. *The sequence $\{A_n(\theta_0)\}$ is an asymptotically most powerful unbiased test of the hypothesis $\theta = \theta_0$, where $A_n(\theta_0)$ denotes the critical region of type A for testing $\theta = \theta_0$.*

ON THE DISTRIBUTION OF THE QUOTIENT OF TWO CHANCE VARIABLES

By J. H. Curtiss

Cornell University

1 **Introduction.** Although the quotient of two chance variables appears frequently in mathematical statistics, the methods used in the literature to derive the distributions of quotients have usually been special ones devised for the particular variables under consideration, and in no way indicative of the general result It is the purpose of this paper to study the distribution of the quotient of two variables for itself alone, with attention first to the question of existence, and then to the accurate derivation of a number of general formulas for the frequency function and d.f.[1] The principal formulas which we shall derive may be described briefly as follows (the numerals refer to the equation numbers in the text).

(3 1). The frequency function of the quotient of two variables which have an absolutely continuous joint probability function.

(4.11), (4.12) The d.f. of the quotient of a pair of arbitrary independent variables, expressed in terms of the d.f.'s of these variables.

(5.2). The d f. of the quotient of a pair of arbitrary independent variables, expressed in terms of the c.f.'s[2] of these variables

(6.4). The limiting form of the d.f. of a quotient of two sums of arbitrary identical independent variables.

(7.1). A formula analogous to (3.1) for the product of two chance variables

(7.2). A formula analogous to (4.11) for the product of two chance variables

2. **The existence of the quotient distribution.** The function $Z = X/Y$ is a continuous function of X and Y, finite and uniquely defined for all points (X, Y) such that $Y \neq 0$. Therefore if $P\{Y = 0\} = 0$, the pr.f.[3] $P(S)$ of the joint distribution of X and Y determines a probability distribution for Z (see [1, pp. 12–13]). To avoid irrelevant difficulties, we shall assume in the sequel that $P\{Y = 0\} = 0$ unless definite statement is made to the contrary. This assumption involves no real restriction on our work, for in situations in which, a priori, the assumption is not fulfilled, we can always replace the distribution

[1] I.e., distribution function The underlying axioms, terminology, and abbreviations in this paper are uniform with those of Cramér's book [1] For the definition of d f., see [1, p 11].

[2] I e., characteristic functions. See [1, p. 23]

[3] I.e., probability function, [1, p 9]

of Y by the conditional distribution of Y relative to the hypothesis that $Y \neq 0$. In such cases, then, the distribution of Z which we are about to study is to be interpreted as a conditional distribution relative to this hypothesis

We shall suppose that the space of X is the x-axis, that of Y, the y-axis, and that of Z, the z-axis. It is quite readily seen that the set of points in the (x, y) plane which corresponds to the set $Z \leqq z$ consists of

(i) the infinite region[4] in the upper half-plane which is bounded by the negative x axis and by the line $x = zy$;

(ii) the infinite region in the lower half-plane bounded by the positive x-axis and the line $x = zy$;

(iii) the line $x = zy$ except for the origin.

Denoting this set by S_z, we have

$$H(z) = \int_{S_z} dP(S) = P(S_z),$$

where $H(z)$ is the d.f. of Z. The present paper, from the viewpoint of analysis, is simply a study of the Lebesgue-Stieltjes integral appearing in this equation.

3. The continuous case. Suppose first that $P(S)$ is absolutely continuous. This means that the joint distribution of X and Y has a frequency function $\varphi(x, y)$, which is defined almost everywhere, is non-negative, and has the property that $P(S) = \int_S \varphi(x, y)\, dx\, dy$. In general, this integral must be taken in the Lebesgue sense, but of course if the discontinuities of φ form a set of two-dimensional measure zero, and if the Jordan content of any bounded portion of the boundary of S is zero, then this integral is just an ordinary improper double Riemann integral.[5] In particular, these conditions are fulfilled if φ is continuous everywhere and if $S = S_z$.

The transformation $x = uv$, $y = v$, gives a continuous one-to-one map of S_z onto a set $\mathfrak{S}_z$ of the (u, v) plane which consists of the closed half-plane lying to the left of the line $u = z$, but with the u-axis deleted. The Jacobian of the transformation has the absolute value $|v|$. By the theorem for change of variables in Lebesgue integrals [4, pp. 653–655], we have

$$H(z) = \int_{S_z} \varphi(x, y)\, dx\, dy = \int_{\mathfrak{S}_z} |v|\, \varphi(uv, v)\, du\, dv.$$

By Fubini's Theorem [6, pp. 203–208], the last integral can be expressed as a repeated integral. Integrating first with respect to v, we obtain this result

THEOREM 3.1: *If the joint variable (X, Y) has the frequency function $\varphi(x, y)$, then*

$$H(z) = \int_{-\infty}^{z} \left[\int_{-\infty}^{+\infty} |v|\, \varphi(uv, v)\, dv \right] du,$$

[4] I e., *open* connected set.

[5] See [4, pp. 476–478: p 575].

and consequently $H(z)$ is an absolutely continuous function of z. The frequency function of the distribution of Z exists almost everywhere, and is given by the formula

$$h(z) = F'(z) = \int_{-\infty}^{+\infty} |v| \varphi(zv, v)\, dv. \tag{3.1}$$

We remark that if X and Y are independent, so that $\varphi(x, y) = f(x) \cdot g(y)$, where f and g are respectively the frequency functions of X and Y, then (3.1) may be written in the form

$$h(z) = \int_{-\infty}^{+\infty} |v| f(zv) g(v)\, dv. \tag{3.2}$$

This case was considered recently by Huntington [5], with the additional restrictions that $g(y) = 0$, $y < 0$, and that $f(x)$ and $g(y)$ be continuous.

All the familiar special quotient distributions of applied mathematical statistics, such as Student's t and Fisher's z, may conveniently and rigorously be derived by means of (3.1) and (3.2); in each case the required result follows immediately after an obvious change of variables in the integrand. We pause here only to point out explicitly the result obtained when X and Y have a normal joint distribution with variances σ_X^2, σ_Y^2, and correlation coefficient ρ. If the means $E(X)$ and $E(Y)$ are not equal to zero, it is apparently impossible to evaluate (3.1) in closed form; this case has been studied in some detail by Geary [3] and by Fieller [2]. But if $E(X) = E(Y) = 0$, then

$$h(z) = \frac{\sigma_X \sigma_Y \sqrt{1 - \rho^2}}{\pi} \cdot \frac{1}{\sigma_Y^2 \left(z - \rho \frac{\sigma_X}{\sigma_Y}\right)^2 + \sigma_X^2 (1 - \rho^2)},$$

which is the frequency function of a Cauchy distribution with mode at the point $z = \rho\sigma_X/\sigma_Y$, the value of the regression coefficient of X on Y. If X and Y are independent, then $\rho = 0$, and the frequency function becomes

$$h(z) = \frac{\sigma_X \sigma_Y}{\pi} \cdot \frac{1}{\sigma_Y^2 z^2 + \sigma_X^2}. \tag{3.3}$$

4. **The quotient of two arbitrary independent variables.** We shall henceforth drop the restriction that $P(S)$ be absolutely continuous, but shall suppose instead that X and Y are independent chance variables with one-dimensional distributions of the most general type, except that the distribution of Y will be subject to the restriction that $P\{Y = 0\} = 0$.

We denote the d.f. of X by $F(x)$, that of Y by $G(y)$, and, as usual, that of Z

by $H(z)$. It is to be noticed that the condition $P\{Y = 0\} = 0$ implies that $G(y)$ is continuous at the point $y = 0$. Let

$$f(t) = \int_{-\infty}^{+\infty} e^{itx}\, dF(x)$$

$$g^+(t) = \int_0^{\infty} e^{ity}\, dG(y) \tag{4.1}$$

$$g^-(t) = \int_{-\infty}^{0} e^{ity}\, dG(y).$$

Clearly

$$H(z) = P\{X - zY \leqq 0;\ Y > 0\} + P\{X - zY \geqq 0;\ Y < 0\}. \tag{4.2}$$

We introduce the functions

$$\Gamma_1(u) = P\{X - zY \leqq u;\ Y > 0\} = [1 - G(0)]\cdot P\{X - zY \leqq u \mid Y > 0\},^6$$

$$\gamma_1(t) = \int_{-\infty}^{+\infty} e^{itu}\, d\Gamma_1(u),$$

$$\Gamma_2(u) = P\{zY - X \leqq u;\ Y < 0\} = G(0)\cdot P\{zY - X \leqq u \mid Y < 0\},$$

$$\gamma_2(t) = \int_{-\infty}^{+\infty} e^{itu}\, d\Gamma_2(u), \tag{4.3}$$

$$\Gamma(u) = \Gamma_1(u) + \Gamma_2(u)$$

$$\gamma(t) = \int_{-\infty}^{+\infty} e^{itu}\, d\Gamma(u) = \gamma_1(t) + \gamma_2(t).$$

By (4.2) and (4.3),

$$H(z) = \Gamma(0). \tag{4.4}$$

We shall now evaluate $\Gamma_1(u)$ and $\Gamma_2(u)$ in terms of $F(x)$ and $G(y)$, and also $\gamma_1(t)$ and $\gamma_2(t)$ in terms of $f(t)$, $g^+(t)$, and $g^-(t)$.

Let us assume for a moment that $P\{Y > 0\} \neq 0$; that is, that $G(0) < 1$. The conditional distribution of Y relative to the hypothesis that $Y > 0$ then has the d.f.

$$G_1(y) = \begin{cases} \dfrac{G(y) - G(0)}{1 - G(0)}, & y \geqq 0, \\ 0, & y < 0. \end{cases} \tag{4.5}$$

The d.f. of $-zY$ relative to this hypothesis is $G_1(-y/z)$ if $z < 0$, and $1 - G_1[(-y/z) - 0]$ if $z > 0$.

---

[6] By $P(A \mid b)$ is meant the conditional probability of the event A relative to the hypothesis b.

It is well known that the corresponding d.f. of the sum $X + (-zY)$ is given by a convolution of the d.f 's of X and $(-zY)$.[7] In the present case, this result takes the form

$$(4\,6)\quad P\{X - zY \leqq u \mid Y > 0\} = \begin{cases} \int_{-\infty}^{+\infty} F(u-v)\, dG_1\left(-\frac{v}{z}\right), & z < 0, \\ \int_{-\infty}^{+\infty} F(u-v)\, d\left[1 - G_1\left(-\frac{v}{z} - 0\right)\right], & z > 0. \end{cases}$$

Referring to the definition of these Lebesgue-Stieltjes integrals [4, pp 662–663], we see that the change of variables $w = -v/z$ yields the equations

$$(4.7)\qquad P\{X - zY \leqq u \mid Y > 0\} = \begin{cases} \int_0^{\infty} F(u + zw)\, dG_1(w), & z < 0, \\ \int_0^{\infty} F(u + zw)\, dG_1(w - 0), & z > 0. \end{cases}$$

Now the definition of the variation of $G_1(y)$ [4, pp 341–342] used in forming these Lebesgue-Stieltjes integrals makes no distinction between the variation of $G_1(y)$ and that of $G_1(y - 0)$ over any bounded set contained in an interval of integration $a \leq y < \infty$, provided that $G_1(y)$ is continuous at a in the two-sided sense. Since $G_1(y)$ is continuous at $y = 0$ in this sense, it is possible to replace $G_1(w - 0)$ by $G_1(w)$ in the second of the two integrals in (4.7).

Equation (4.7) is clearly true for $z = 0$ as well as for all other values of z. Referring to (4 5) and (4.3), we see that

$$\Gamma_1(u) = \int_0^{\infty} F(u + zw)\, dG(w), \quad \text{all } z.$$

The c.f of the convolution (4 6) is the product of the c.f.'s of X and of the conditional distribution of $-zY$ [1, p 36] This product is $f(t) \int_0^{\infty} e^{-itzv}\, dG_1(y)$. Thus by (4.5), (4.3), and (4.1),

$$(4.8)\qquad \gamma_1(t) = [1 - G(0)]\left[f(t) \cdot \int_0^{\infty} e^{-itzv}\, dG_1(y)\right] = f(t)g^+(-tz).$$

We have established (4 7) and (4.8) under the condition that $P\{Y > 0\} \neq 0$. However, it is obvious that they are trivially true if $P\{Y > 0\} = 0$.

We turn now to $\Gamma_2(u)$. Supposing that $P\{Y < 0\} \neq 0$, the conditional distribution of Y relative to the hypothesis that $Y < 0$, has the d.f.

$$G_2(y) = \begin{cases} \frac{G(y)}{G(0)}, & y < 0, \\ 1, & y \geqq 0. \end{cases}$$

[7] See [1, pp 35–36]; also [7].

The conditional distribution of zY has the d.f. $G_2(y/z)$ for $z > 0$, and $1 - G_2[(y/z) - 0]$ for $z < 0$. The d.f. of $-X$ is $1 - F(-x - 0)$. Thus

$$P\{zY - X \leqq u \mid Y < 0\} =$$

$$\begin{cases} \int_{-\infty}^{+\infty} \{1 - F[-(u-v) - 0]\}\, d\left[1 - G_2\left(\frac{v}{z} - 0\right)\right], & z < 0, \\ \int_{-\infty}^{+\infty} \{1 - F[-(u-v) - 0]\}\, dG\left(\frac{v}{z}\right), & z > 0, \end{cases}$$

$$= 1 - \int_{-\infty}^{0} F(zw - u - 0)\, dG_2(w).$$

Evidently the first and last members of this equation are equal for $z = 0$ as well as for all other values of z. From (4.3) we obtain

$$\Gamma_2(u) = G(0) - \int_{-\infty}^{0} F(zw - u - 0)\, dG(w), \quad \text{all } z.$$

Also, as before,

$$\gamma_2(t) = f(-t)g^-(zt).$$

Obviously, the last two equations are still true if $P\{Y < 0\} = 0$

To summarize, we have shown that

$$\Gamma(u) = G(0) + \int_0^{\infty} F(u + zw)\, dG(w) - \int_{-\infty}^{0} F(zw - u - 0)\, dG(w), \quad \text{all } z; \tag{4.9}$$

$$\gamma(t) = f(t)g^+(-zt) + f(-t)g^-(zt). \tag{4.10}$$

Referring now to (4.4) and letting $u = 0$ in (4.9), we are able to state the following theorem:

THEOREM 4.1: *If X and Y are independent chance variables with respective d.f.'s $F(x)$ and $G(y)$, the d.f. of the quotient X/Y is given by the formula*

$$H(z) = G(0) + \int_0^{\infty} F(zw)\, dG(w) - \int_{-\infty}^{0} F(zw - 0)\, dG(w) \tag{4.11}$$

for all values of z.

We shall not attempt to make a careful study of the above formula, such as the studies which certain writers have made of convolutions. However, it does seem desirable to place on record here certain remarks concerning it of a more or less superficial character. For convenience in later reference, we state these remarks in the form of four lemmas.

LEMMA 4.1. *Let M_1 be the set of all values of z such that if $z \in M_1$, the set of discontinuity points of $F(zw)$ on the w-axis has a point in common with the point spectrum of $G(w)$. Then if $z \in C(M_1)$,[8] the integrals $\int_0^{\infty} F(zw \pm 0)\, dG(w)$,*

[8] By $C(M_1)$ we mean the complement of M_1 with respect to the z-axis.

$\int_{-\infty}^{0} F(zw \pm 0)\, dG(w)$, *are Riemann-Stieltjes integrals and consequently the integrands can be replaced by* $F(zw)$ *without altering the values of the integrals*

The lemma follows immediately from the definitions of Riemann-Stieltjes and Lebesgue-Stieltjes integrals

LEMMA 4.2: *The set* M_1 *is denumerable.*

The proof can easily be supplied by the reader

LEMMA 4.3: *Let* M_2 *be the set of all values of* z *such that if* $z \in M_2$, $\Gamma(u)$ *is discontinuous at* $u = 0$. *Then* $M_2 \subset M_1$.

To prove this statement, we first observe that $\Gamma(u)$ is a genuine d.f. [1, p. 11]. For obviously $\Gamma(-\infty) = 0$, $\Gamma(+\infty) = 1$, and since $\Gamma_1(u)$ and $\Gamma_2(u)$ are both products of d.f.'s into constants, these two functions, and therefore $\Gamma(u)$, must be continuous from the right It is this last property of $\Gamma(u)$ which is needed for our present purposes; in particular, we have the relation $\lim_{u\to+0}\Gamma(u) = \Gamma(0) = H(z)$. On the other hand, by the general convergence theorem for Lebesgue-Stieltjes integrals [4, pp. 663–664], we have

$$\lim_{u\to-0} \Gamma(u) = G(0) + \int_0^\infty F(zw - 0)\, dG(w) - \int_{-\infty}^{0} F(zw)\, dG(w).$$

If z be chosen so that this integral and the ones in (4 11) are all Riemann-Stieltjes integrals, the expression $(zw - 0)$, wherever it appears, may be replaced by zw without changing the values of the integrals. Thus for such a value of z, $\Gamma(+0) = \Gamma(-0)$. According to Lemma 4 1, we can be sure that at least if $z \in C(M_1)$, the integrals here will be Riemann-Stieltjes integrals, so our proposition is proved

Since $H(z_1 + 0)$ is equal to $\Gamma(+0)$ with $z = z_1$, and $H(z_1 - 0)$ is equal to $\Gamma(-0)$ with $z = z_1$, we have the following result:

LEMMA 4.4. *The set* M_2 *is the set of discontinuity points of* $H(z)$.

By using the alternate form of the convolutions used to derive (4 9), we obtain a representation of $\Gamma(u)$ somewhat more complicated than that appearing in (4.9). The corresponding formula for $H(z)$ is as follows:

$$(4.12)\quad H(z) = \begin{cases} G(0)[1 - F(-0)] - G(0)F(0) + \int_{-\infty}^{0} G\left(\frac{v}{z}\right) dF(v) \\ \qquad - \int_0^\infty G\left(\frac{v}{z} - 0\right) dF(v - 0), & z < 0; \\ F(0)[1 - G(0)] + G(0)[1 - F(-0)], & z = 0; \\ 1 + G(0)[1 - F(-0)] - G(0)F(0) + \int_{-\infty}^{0} G\left(\frac{v}{z}\right) dF(v - 0) \\ \qquad - \int_0^\infty G\left(\frac{v}{z} - 0\right) dF(v), & z > 0 \end{cases}$$

5. **Representation of $H(z)$ by characteristic functions.** A simple algebraic formula connecting the c.f. of Z with those of X and Y is not available. However, there exists an interesting representation of $H(z)$ in terms of the functions $f(t)$, $g^+(t)$, and $g^-(t)$. The result may be stated as follows:

THEOREM 5.1:[9] *Let the distributions of the independent variables X and Y have finite first absolute moments, and let the integral*

$$\left(\int_{-\infty}^{-1} + \int_{1}^{\infty}\right) \left| f(t)g^+(-zt) + \frac{f(-t)g^-(zt)}{t} \right| dt \tag{5.1}$$

be finite for each value of z. Let $\Delta(u)$ be any d.f. with a finite first absolute moment, and let $\left(\int_{-\infty}^{-1} + \int_{1}^{\infty}\right) \left| \frac{\delta(t)}{t} \right| dt$ be finite, where $\delta(t)$ is the c.f. of $\Delta(u)$. Then

$$H(z) = \Delta(0) - \frac{1}{2\pi i} \int_{-\infty}^{+\infty} \frac{f(t)g^+(-zt) + f(-t)g^-(zt) - \delta(t)}{t}\, dt. \tag{5.2}$$

If the integral obtained by formal differentiation under the integral sign with respect to z in (5.2) is uniformly convergent in a certain interval I, then the frequency function $h(z)$ of the distribution of z exists in that interval and is given by the formula

$$h(z) = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} [f(t)g^{+\prime}(-zt) - f(-t)g^{-\prime}(zt)]\, dt, \qquad z \in I.$$

We remark that the condition (5.1) will be satisfied for all values of z if $f(t)$ alone satisfies a similar condition, inasmuch as $|g^+(t)| \leqq 1$, $|g^-(t)| \leqq 1$. Important special cases of the theorem arise when $\Delta(u)$ is replaced by $F(u)$ or $G(u)$, and when $\Delta(u)$ is so chosen that $\Delta(0) = 0$.

Our proof of the theorem will depend on a rather general result due to Cramér [1, Theorem 12], which we shall restate here in the special form applicable to the problem at hand.

LEMMA 5.1: *Let $R(u)$ be a function of bounded variation over the infinite interval $-\infty < u < \infty$, let $\lim_{u \to -\infty} R(u) = \lim_{u \to +\infty} R(u) = 0$, and let $r(t) = \int_{-\infty}^{+\infty} e^{itu}\, dR(u)$. If (a) $\int_{-\infty}^{+\infty} |u|\, dR(u)$ and (b) $\left(\int_{-\infty}^{-1} + \int_{1}^{\infty}\right) \left| \frac{r(t)}{t} \right| dt$, both are finite, then for every value of u,*

$$R(u) = -\frac{1}{2\pi i} \int_{-\infty}^{+\infty} \frac{r(t)}{t}\, e^{-itu}\, dt.$$

To prove Theorem 5.1, we observe that since $\Gamma(u)$ is a d.f. (see proof of Lemma 4.3), the difference $\Gamma(u) - \Delta(u)$ is a function similar to the function $R(u)$ of the lemma. If we do let $R(u) = \Gamma(u) - \Delta(u)$, it follows at once that $r(t) = \gamma(t) - \delta(t) = f(t) \cdot g^+(-zt) + f(-t)g^-(zt) - \delta(t)$. If we can verify that this $R(u)$

[9] The theorem is due to Cramér in the case in which $G(0) = 0$, and $\Delta(u) = G(u)$. See [1, Theorem 16].

satisfies conditions (a) and (b) of the lemma, then we shall have established the relation,

$$\Gamma(u) = \Delta(u) - \frac{1}{2\pi i}\int_{-\infty}^{+\infty} \frac{f(t)g^+(-zt) + f(-t)g^-(zt) - \delta(t)}{t} e^{-iut}\, dt,$$

for all values of u, and letting $u = 0$ in this equation, we shall obtain (5.2).

Condition (b) in the lemma is taken care of by (5.1) and the condition on $\delta(t)$ in Theorem 5 1. Clearly condition (a) will be satisfied if it turns out that $\Gamma(u)$ has a finite first absolute moment. Now the existence of finite first absolute moments of X and Y will insure the existence of finite first absolute moments for the conditional distributions involved in the definitions of $\Gamma_1(u)$ and $\Gamma_2(u)$, because $E\,|\,X - zY\,| \leqq E\,|\,X\,| + |\,z\,|\,E\,|\,Y\,|$ It follows quite readily from this that the first absolute moment of $\Gamma(u)$ is finite. The proof of the theorem is complete.

6. **Distributions of variable form.** We consider now the case in which the distributions of the numerator and denominator approach limiting forms.

THEOREM 6 1: *Let the independent variables X_α and Y_β have respective d f.'s $F_\alpha(x)$ and $G_\beta(y)$ which depend upon the two parameters α and β. Let $H_{\alpha,\beta}(z)$ be the d f. of the quotient $Z_{\alpha,\beta} = X_\alpha/Y_\beta$. If there exist two chance variables X and Y with respective distribution functions $F(x)$ and $G(y)$ such that $\lim_{\alpha\to\infty} F_\alpha(x) = F(x)$ at all points of continuity of $F(x)$, and $\lim_{\beta\to\infty} G_\beta(y) = G(y)$, at all points of continuity of $G(y)$, then*

$$\lim_{\substack{\alpha\to\infty\\ \beta\to\infty}} H_{\alpha,\beta}(z) = \lim_{\alpha\to\infty}\lim_{\beta\to\infty} H_{\alpha,\beta}(z) = \lim_{\beta\to\infty}\lim_{\alpha\to\infty} H_{\alpha,\beta}(z) = H(z) \tag{6.1}$$

at all points of continuity of $H(z)$, where $H(z)$ is the d f of the variable X/Y The double limit in (6.1) is uniform in any finite or infinite interval of continuity of $H(z)$

In the interpretation of the limits involved in this theorem, it is to be understood that in the hypotheses, α may tend to infinity over any unbounded set T_α of the α-axis, and β may tend to infinity over any unbounded set T_β of the β-axis, provided that in (6.1), α and β are restricted so that $\alpha \in T_\alpha$ and $\beta \in T_\beta$.

To prove the theorem, we introduce functions $f_\alpha(t)$, $g_\beta^+(t)$, $g_\beta^-(t)$, $\Gamma_{\alpha,\beta}(u)$, $\gamma_{\alpha,\beta}(t)$, which are defined by equations (4.1) and (4.3) with F, G, X, Y replaced respectively by F_α, G_β, X_α, Y_β. On the other hand, with reference to the distributions of X and Y, we employ the notation of section 4 without modification. According to the work in that section, $\Gamma(u)$ is given by (4.9) and its c f. $\gamma(t)$ is given by (4.10). Also,

$$\gamma_{\alpha,\beta}(t) = f_\alpha(t)g_\beta^+(-zt) + f_\alpha(-t)g_\beta^-(zt).$$

But it is an immediate consequence of our hypotheses that $\lim_{\alpha\to\infty} f_\alpha(t) = f(t)$,

$\lim_{\beta\to\infty} g_\beta^+(t) = g^+(t)$, and $\lim_{\beta\to\infty} g_\beta^-(t) = g^-(t)$, all of the limits being uniform in any finite interval of values of t.[10] Thus

$$\lim_{\substack{\alpha\to\infty\\ \beta\to\infty}} \gamma_{\alpha,\beta}(t) = \lim_{\alpha\to\infty}\lim_{\beta\to\infty} \gamma_{\alpha,\beta}(t) = \lim_{\beta\to\infty}\lim_{\alpha\to\infty} \gamma_{\alpha,\beta}(t) = \gamma(t), \tag{6.2}$$

uniformly in any finite interval on the t-axis.

Consider the extreme members of (6.2). It follows immediately from a well-known general theorem[11] that $\lim_{\alpha\to\infty,\beta\to\infty} \Gamma_{\alpha,\beta}(u) = \Gamma(u)$ at all continuity points of $\Gamma(u)$. Then since $H_{\alpha,\beta}(z) = \Gamma_{\alpha,\beta}(0)$ and $H(z) = \Gamma(0)$, we find that

$$\lim_{\substack{\alpha\to\infty\\ \beta\to\infty}} H_{\alpha,\beta}(z) = H(z), \qquad z \in C(M_2),$$

where M_2 is the set defined in Lemma 4.3. By Lemma 4.4, the set M_2 is the set of discontinuity points of $H(z)$, so the equality of the first and last members of (6.1) is established at all continuity points of $H(z)$. The uniformity of the limit is due to a general property of convergent sequences of d.f.'s, see [1, p. 31].

The existence and equivalence to $H(z)$ of each of the iterated limits in (6 1) may be established by two consecutive applications of the foregoing argument, and by the use of (6.2). We leave the details to the reader.

It is to be remarked that both $H_{\alpha,\beta}(z)$ and $H(z)$ can be represented by (4 11), provided, of course, that F and G in (4.11) are replaced by F_α and G_β in the case of $H_{\alpha,\beta}$; thus our theorem essentially states that the order of the double limit and the integration is immaterial in this formula. A similar remark applies to formula (5.2).

The reader is reminded that we have tacitly been assuming that the d.f. of any variable appearing in a denominator is continuous at the origin. In case $G_\beta(y)$ does not satisfy this condition, but $G(y)$ does satisfy it, and if, as suggested in section 2, we consider $H_{\alpha,\beta}(y)$ to be the d.f. of the conditional distribution of $Z_{\alpha,\beta}$ relative to the hypothesis that $Y_\beta \neq 0$, then it can be shown rather easily that Theorem 6.1 remains true with this modified interpretation. But if $G(y)$ is discontinuous at the origin, and if $H(z)$ is interpreted as the d.f. of the conditional distribution, then (6.1) may be no longer true, as can be shown by trivial examples.

Perhaps the most important cases of variable distributions arise in the consideration of sums of independent chance variables. We accordingly present the following synthesis of Theorem 6.1 and a simple case of the Central Limit Theorem.

THEOREM 6.2: *Let $U_1, U_2, \cdots$, be a sequence of identically distributed chance variables, each with mean zero and (finite) standard deviation σ_U, and let V_1,*

[10] See [1, p. 30].

[11] See [1, Theorem 11]. The result needed here is a trivial extension of the theorem cited.

$V_2, \cdots$, be a sequence of identically distributed chance variables, each with mean zero and (finite) standard deviation σ_V. Furthermore, let the variables U_i and V_j be all independent, $i = 1, 2, \cdots, j = 1, 2, \cdots$. If m and n tend to infinity in such a way that

$$\lim_{\substack{m\to\infty \\ n\to\infty}} \sqrt{\frac{n}{m}} = k \neq 0, \tag{6.3}$$

then the d.f. of the conditional distribution of the variable

$$W_{m,n} = \frac{U_1 + U_2 + \cdots + U_m}{V_1 + V_2 + \cdots + V_n},$$

relative to the hypothesis that the denominator is different from zero, tends uniformly to the function

$$J(w) = \int_{-\infty}^{w} \frac{k\sigma_U\sigma_V}{\pi} \cdot \frac{1}{\sigma_V^2 k^2 u^2 + \sigma_U^2}\, du. \tag{6.4}$$

For if we let

$$Z_{m,n} = \frac{\dfrac{U_1 + U_2 + \cdots + U_m}{\sigma_U\sqrt{m}}}{\dfrac{V_1 + V_2 + \cdots + V_n}{\sigma_V\sqrt{n}}},$$

then $W_{m,n} = \sqrt{m/n}(\sigma_U/\sigma_V)Z_{m,n}$ The Central Limit Theorem [1, Theorem 20] states that the d.f.'s of the numerator and denominator of $Z_{m,n}$ each tend to the function $\int_{-\infty}^{x} (1/\sqrt{2\pi})e^{-t^2/2}\, dt$, which is the d.f. of a normal distribution with mean zero and variance one. By (3 3), the quotient of two variables, each of which has this d.f , has the continuous d f. $H(z) = \int_{-\infty}^{z} (1/\pi)[1/(1 + x^2)]\, dx$. If we let $H_{m,n}(z)$ denote the d.f of the conditional distribution of $Z_{m,n}$, relative to the hypothesis that the denominator of $Z_{m,n}$ is different from zero, then by Theorem 6.1, $\lim_{m\to\infty, n\to\infty} H_{m,n}(z) = H(z)$ uniformly in z. Now the d.f. of the conditional distribution of $W_{m,n}$ is $H_{m,n}[\sqrt{n/m}(\sigma_V/\sigma_U)w]$, and because of (6.3) and the uniformity of the limit of $H_{m,n}(z)$, this approaches $H[k(\sigma_V/\sigma_U)w]$. Differentiating the last expression with respect to w, we find that the resulting frequency function is equal to $J'(w)$; and this concludes the proof.

As an application of the theorem, let us consider the following problem. From an urn containing white and black balls in the proportion of p to $1 - p$, we shall make 100 random drawings of a single ball with replacement after each drawing. Let $W_{50,50}$ be the ratio of the deviation of the number of white balls in the first 50 drawings from the expected number, to the deviation of the number of white balls in the second 50 drawings from the expected number. What is

the approximate value of w for which $P\{W_{50,50} \geqq w \mid b\} = .05$, where the hypothesis b is that the denominator of $W_{50,50}$ shall be different from zero?[12]

To answer this question, we observe that the numerator and denominator of $W_{50,50}$ can each be expressed as the sum of 50 independent identical chance variables, each with mean zero and with variance $p(1 - p)$. Thus according to Theorem 6.2, the approximate d.f. of $W_{50,50}$ is

$$J(w) = \int_{-\infty}^{w} \frac{1}{\pi} \frac{1}{1 + u^2} du = \frac{1}{2} + \frac{1}{\pi} \arctan w,$$

and the required value of w satisfies the equation $J(\infty) - J(w) = .05$. The solution of this equation (correct to one decimal place) is $w = 6.3$

It is perhaps needless to remark that a study of the error involved in supposing $J(w)$ to be the d. f. of $W_{m,n}$ in Theorem 6.2, must necessarily precede the unreserved acceptance of numerical results obtained by means of that theorem.

7. **Products of chance variables.** We conclude this paper with a rather brief treatment of the distribution of the product of two chance variables. To preserve a notation uniform with that of the preceding sections, we shall write the product as $X = YZ$, where the d f.'s of X, Y, and Z are to be denoted, as before, by $F(x)$, $G(y)$, and $H(z)$, respectively. The existence of $F(x)$ is readily proved by the methods of section 2. The assumption that $P\{Y = 0\} = 0$ is of course unnecessary here, and will be dropped in this section

In the continuous case, an argument similar to the one employed in section 3 will establish the following result:

THEOREM 7.1: *If the joint variable (Y, Z) has the frequency function $\psi(y, z)$, then*

$$F(x) = \int_{-\infty}^{x} \left[\int_{-\infty}^{+\infty} \left| \frac{1}{v} \right| \psi\left(\frac{u}{v}, v\right) dv \right] du$$

$$= \int_{-\infty}^{x} \left[\int_{-\infty}^{+\infty} \left| \frac{1}{v} \right| \psi\left(v, \frac{u}{v}\right) dv \right] du,$$

and consequently $F(x)$ is an absolutely continuous function of x. The frequency function of the distribution of X exists almost everywhere, and is given by the formula

$$(7.1) \qquad f(x) = F'(x) = \int_{-\infty}^{+\infty} \left| \frac{1}{v} \right| \psi\left(\frac{x}{v}, v\right) dv = \int_{-\infty}^{+\infty} \left| \frac{1}{v} \right| \frac{\psi}{v}\left(v, \frac{x}{v}\right) dv.$$

In the discontinuous case, with Y and Z independent, we can write $X = ZY = Z/(1/Y)$ and use Theorem 4.1 to derive a formula for $F(x)$. We have:

$$F(x) = P\{X \leqq x\} = P\{Y \neq 0\} P\{X \leqq x \mid Y \neq 0\} + P\{X \leqq x; Y = 0\}.$$

[12] This hypothesis would always be fulfilled in case $50p$ is not an integer.

Excluding for a moment the trivial case in which $P\{Y \neq 0\} = 0$, let $G_1(y)$ be the d.f. of the conditional distribution of $(1/Y)$ relative to the hypothesis that $Y \neq 0$ Then

$$P\{Y \neq 0\}G_1(y) = \begin{cases} G(-0) + 1 - G\left(\frac{1}{y} - 0\right), & y > 0, \\ G(-0), & y = 0, \\ G(-0) - G\left(\frac{1}{y} - 0\right), & y < 0. \end{cases}$$

It is to be observed that $G_1(y)$ is continuous at $y = 0$ Using Theorem 4.1, we find that

$$P\{X \leqq x \mid Y \neq 0\} = G_1(0) + \int_0^\infty H(xw)\, dG_1(w) - \int_{-\infty}^0 H(xw - 0)\, dG_1(w).$$

So

$$\begin{aligned} &P\{Y \neq 0\}P\{X \leqq x \mid Y \neq 0\} \\ &\quad = G(-0) + \int_{0+0}^\infty H(xw)\, d\left[-G\left(\frac{1}{w} - 0\right)\right] - \int_{-\infty}^{0-0} H(xw - 0)\, d\left[-G\left(\frac{1}{w} - 0\right)\right] \\ &\quad = G(-0) + \int_{0+0}^\infty H\left(\frac{x}{v}\right) dG(v) - \int_{-\infty}^{0-0} H\left(\frac{x}{v} - 0\right) dG(v). \end{aligned}$$

This equation is trivially true if $P\{Y \neq 0\} = 0$. Also,

$$P\{X \leqq x; Y = 0\} = \begin{cases} 0, & x < 0, \\ G(0) - G(-0), & x \geqq 0. \end{cases}$$

Thus we obtain the following theorem:

THEOREM 7.2 *If Y and Z are independent chance variables with respective d.f.'s $G(y)$ and $H(z)$, then the d.f. of their product is given by the formula*

$$(7.2) \qquad F(x) = \int_{0+0}^\infty H\left(\frac{x}{v}\right) dG(v) - \int_{-\infty}^{0-0} H\left(\frac{x}{v} - 0\right) dG(v) + \begin{cases} G(-0), & x < 0, \\ G(0), & x \geqq 0, \end{cases}$$

for all values of x.

REFERENCES

[1] H. CRAMÉR, *Random Variables and Probability Distributions*, Cambridge, 1937.

[2] E. C FIELLER, "The distribution of the index in a normal bivariate population," *Biometrika*, Vol. 24 (1932), pp 428–440

[3] R. C. GEARY, "The frequency distribution of the quotient of two normal variates," *Roy Stat. Soc Jour.*, Vol. 93 (1930), pp. 442–446

[4] E. W. HOBSON, *The Theory of Functions of a Real Variable*, Vol. 1, Cambridge, 1927.

[5] E. V. HUNTINGTON, "Frequency distribution of product and quotient," *Annals of Math. Stat.*, Vol. 10 (1939), pp 195–198.

[6] H KESTELMAN, *Modern Theories of Integration*, Oxford, 1937

[7] A. WINTNER, "On the addition of independent distributions," *Am. Jour Math.*, Vol. 56 (1934), pp. 8–16.

SOME GENERALIZATIONS OF THE LOGARITHMIC MEAN AND OF SIMILAR MEANS OF TWO VARIATES WHICH BECOME INDETERMINATE WHEN THE TWO VARIATES ARE EQUAL

By Edward L. Dodd

University of Texas

1. Introduction. The logarithmic mean m of positive numbers, x and y, as given by

$$m = \frac{y - x}{\log_e y - \log_e x} = \frac{y - x}{\log_e (y/x)} \tag{1}$$

is of considerable importance in problems[1] relating to the flow of heat.

The logarithmic mean arises, moreover, in less technical problems such as the following: Given that incomes t in the interval, $x \leqq t \leqq y$, are distributed with frequency inversely proportional to t. That is, with $k = a$ positive constant,

$$\phi(t)\, dt = (k/t)\, dt \tag{2}$$

is the number of individuals with incomes lying between t and $t + dt$ Then, with $x > 0$, the total number f of individual incomes is

$$f = \int_x^y \phi(t)\, dt = k(\log y - \log x). \tag{3}$$

The combined income g of the group is

$$g = \int_x^y t\phi(t)\, dt = k(y - x). \tag{4}$$

And thus the *logarithmic* mean g/f of the *two* numbers x and y in (1) is the *arithmetic mean of all* the incomes; that is, the *average income*—at least to a close approximation if the group is large enough that integration may replace summation.

Now m in (1) becomes *indeterminate*, if $x = y$. Nevertheless, if $c > 0$, and $x \to c$ and $y \to c$, then $m \to c$. Thus, we may properly speak of m as a mean of these two variates, x and y.

This logarithmic mean is one of a set of means studied by Renzo Cisbani[2], the general form being

---

[1] See Walker, Lewis, and McAdams, *Principles of Chemical Engineering*, McGraw Hill & Co., Part IV, Logarithmic mean temperature difference.

[2] R. Cisbani, "Contributi alla teoria delle medie." *Metron*, Vol. 13(1938), pp. 23-34.

$$z = \left[\frac{b^{x+j} - a^{x+j}}{(x/j + 1)(b^j - a^j)} \right]^{1/x} \tag{5}$$

and the logarithmic mean appearing when $x = 1, j \to 0$

In a chart between pages 28 and 29 Cisbani exhibits thirty varieties of these means (5) It will be noticed that z is *indeterminate* if $a = b$

Some methods for dealing with means which may become indeterminate forms I have indicated in a recent paper.[3]

Now a generalization from a mean for *two* variates to a mean of three or more variates may sometimes *seem* to be *immediate*. However, for the arithmetic mean $(x + y)/2$ of two variates x and y, the function [min. (x, y, z) + max. (x, y, z)]/2 is as much a generalization as is the arithmetic mean $(x + y + z)/3$. *Actually*, the direction in which generalization is to take place is *arbitrary*. However, it is natural to expect the generalization to arise from a problem somewhat similar to one that may give rise to the original mean And it is desirable that to the generalization should be carried over as many properties or characteristics of the original as is possible

In the foregoing illustration, we considered a *single* interval $x \leqq t \leqq y$ in which incomes are distributed in accordance with a *relative* frequency proportional to $\phi(t)$. And the *arithmetic* mean of *all* these incomes was obtained as a *logarithmic* mean of the *two* range limits x and y, at least approximately, allowing integration to take the place of summation If $\phi(t)$ had been $kt^{-3/2}$, instead of kt^{-1}, then the average of *all* the incomes would have been the *geometric* mean of the *two* range limits x and y.

To effect a *first* generalization, we shall now suppose an original interval x_0 to x_n, to be divided into n subintervals by points x_r such that

$$x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n . \tag{6}$$

For each subinterval x_{r-1} to x_r the *same* function $\phi(t)$ will be used to describe the *relative* frequency; but the *total population* for this subinterval will be controlled by a positive constant k_r, in general *different* for the different subintervals. This may be described as *stratification*. To make this more concrete, let us suppose, as before, that $\phi(t) = k/t$ Then, with $x_0 > 0$, the mean M, which will be described more in detail in the next section, will take the form

$$M = \frac{\sum_1^n k_r(x_r - x_{r-1})}{\sum_1^n k_r \log (x_r/x_{r-1})} . \tag{7}$$

Applied to incomes, M would, like m in (1), give average income. To get some idea of the significance of k_r, let us imagine that in some community there are f_r individuals in the income bracket x_{r-1} to x_r, say from \$1001 to \$2000. Let us suppose now that f_r other individuals with incomes between \$1001 and \$2000 distributed in exactly the same manner move into this same community.

[3] "The substitutive mean and certain subclasses of this general mean" *Annals of Math. Stat.*, Vol. 11(1940), pp 163-176 See p 171.

Then k_r would be changed to $k_r' = 2k_r$. But, of course, among the entire $2f_r$ individuals the *relative* distribution of incomes is *exactly* the *same* as among the original f_r individuals.

In *this interpretation* k_r is a *weight* for a *bracket of items*. But, taking M in (7) just as it stands, k_r is the *weight* for the *consecutive pair* of *numbers* x_{r-1} and x_r.

2. The first generalization. When t is in some interval, $I = (a, a')$, finite or infinite, let $\phi(t)$ be a non-negative, integrable function of t.

And in I let the points at which $\phi(t) = 0$, if any, form a null-set. Then, with t in I, write

$$\Phi(t) = \int_a^t \phi(t)\, dt. \tag{8}$$

And, supposing that in (6), $a < x_0$, $a_n < a'$, set

$$f_r = \int_{x_{r-1}}^{x_r} \phi(t)\, dt = \Phi(x_r) - \Phi(x_{r-1}); \qquad r = 1, 2, \cdots, n. \tag{9}$$

Then $f_r > 0$; since $\phi(t) > 0$ and is continuous almost everywhere in (x_{r-1}, x_r). Since in any finite subinterval of I, $t\phi(t)$ is integrable, we may set

$$\Psi(t) = \int_a^t \psi(t)\, dt = \int_a^t t\phi(t)\, dt. \tag{10}$$

$$g_r = \int_{x_{r-1}}^{x_r} \psi(t)\, dt = \Psi(x_r) - \Psi(x_{r-1}). \tag{11}$$

Now, by a mean value theorem, there exists a number t_r' such that

$$g_r/f_r = t_r', \qquad x_{r-1} < t_r' < x_r. \tag{12}$$

Taking *positive* numbers k_r, the weighted arithmetic mean of g_r/f_r, with weights $k_r f_r$ is then

$$M = \frac{\sum_1^n k_r g_r}{\sum_1^n k_r f_r} = \frac{\sum_1^n k_r[\Psi(x_r) - \Psi(x_{r-1})]}{\sum_1^n k_r[\Phi(x_r) - \Phi(x_{r-1})]}. \tag{13}$$

If $\phi(t) = k/t$, this becomes the mean (7) associated with the logarithmic mean. Now, since for (13) the weights $k_r f_r$ are *positive*, it follows from (12) that

$$x_0 < t_1' \leqq M \leqq t_n' < x_n. \tag{14}$$

Suppose, now, that b lies in I, and that subject to (6) each $x_r \to b$. Then, by (14), $M \to b$. And thus M is an *internal mean* of $x_0, x_1, \cdots, x_n$, although with the x's all equal, M assumes an indeterminate form.

In (13) the *weights* k_r are applied to *pairs* of numbers, either to $\Psi(x_r) - \Psi(x_{r-1})$ or to $\Phi(x_r) - \Phi(x_{r-1})$, whereas in *most* weighted means, the weights are applied

to *individual* numbers. We consider now a form equivalent to (13), but in which the weights c_r are attached to the *individual* numbers. It seemed possible to get a more general mean than (13) by abandoning certain conditions upon the weights c_r which first arose. But such relaxing of restrictions leads to difficulties, as will be shown. By setting

$$c_0 = -k_1, \qquad c_n = k_n; \qquad c_r = k_r - k_{r+1}, \qquad r = 1, 2, \cdots, n-1, \tag{15}$$

we may write M in the form;

$$M = \frac{\sum_0^n c_r \Psi(x_r)}{\sum_0^n c_r \Phi(x_r)}. \tag{16}$$

On the other hand, if we choose c's subject to

$$c_0 < 0, \qquad c_r < -(c_0 + c_1 + \cdots + c_{r-1}) \quad \text{for } 0 < r < n, \tag{17}$$

$$c_n = -\sum_0^{n-1} c_r; \tag{18}$$

then positive k's can be found to pass from (16) back to (13).

The question arises whether if the conditions (17) are abandoned, and with the c_r not all zero, (18) is retained as

$$\sum_0^n c_r = 0; \qquad \text{Some } c_r \neq 0, \tag{19}$$

M in (16) will continue to be a mean of $x_0, x_1, \cdots, x_n$, possibly, an external mean.

It may be noted that the condition $\sum c_r = 0$ arises from the fact that when parentheses are removed from (13), each k_r is matched by $-k_r$.

By an example, it will be shown that under (19) alone, M in (16) may fail to be a mean. In (8) and (10) take $a = 0$. Then with $n = 2$, $\phi(t) = t$, take $c_0 = 1$, $c_1 = -2$, $c_2 = 1$ in (16). Then

$$M = \frac{x_0^2 - 2x_1^2 + x_2^2}{2(x_0 - 2x_1 + x_2)}. \tag{20}$$

If $b > 0$, $\epsilon = x_0 - b$, $\eta = x_1 - b$, and $\xi = x_2 - b$, then

$$M = b + \frac{1}{2}\,\frac{\epsilon^2 - 2\eta^2 + \xi^2}{\epsilon - 2\eta + \xi}. \tag{21}$$

If now $\eta = 2\epsilon$, and $\xi = 3\epsilon + \epsilon^2$, then

$$M = b + (2 + 6\epsilon + \epsilon^2)/2 \to b + 1, \qquad \text{as } \epsilon \to 0. \tag{22}$$

Since M does not approach b here, when x_0, x_1, and $x_2 \to b$, in the manner specified, M in (20) is *not* a mean of x_0, x_1, and x_2.

We may enquire, further, whether the function M in (16) could be a mean if, discarding (13), (17) and (18), we put upon c_r the single restriction $c_r > 0$. In that case, if $x_0 < t < x_n$, then, since $\Phi(t)$ and $\Psi(t)$ are continuous functions of t—see (8), (10)—it would follow that if each $x_r \to t$, then $M \to \Psi(t)/\Phi(t)$. But

if M is to be a mean of $x_0, x_1, \cdots, x_n$, then $M \to t$ when each $x_r \to t$. Thus we are led to $\Psi(t) = t\Phi(t)$. Except possibly for points of a null set, $\Phi(t)$ and $\Psi(t)$ have derivatives $\phi(t)$ and $\psi(t)$; and thus

$$\psi(t) = \Psi'(t) = t\Phi'(t) + \Phi(t) = t\phi(t) + \Phi(t). \tag{23}$$

But then, since $\psi(t) = t\phi(t)$ [see (10)] it would follow that $\Phi(t) = 0$ almost everywhere in I; but $\Phi(t) > 0$, if $t > a$. Hence the assumption $c_r > 0$ is not sufficient to make the function in (16) a mean of $x_0, x_1, \cdots, x_n$.

In the simple case of $n = 1$, M becomes

$$M = \frac{\Psi(x_1) - \Psi(x_0)}{\Phi(x_1) - \Phi(x_0)}; \tag{24}$$

and this is a *symmetrical* function of x_0 and x_1.

The question arises whether if $n > 1$, M in (13) or (16) can be a symmetrical function of $x_0, x_1, \cdots, x_n$. Assume, if possible, that with $x < y < z$,

$$H(x, y, z) = \frac{c_0\Psi(x) + c_1\Psi(y) + c_2\Psi(z)}{c_0\Phi(x) + c_1\Phi(y) + c_2\Phi(z)} \tag{25}$$

is a symmetrical function of x, y and z. Now if $a/b = c/d$, and $b - d \neq 0$, it is well known that $a/b = (a - c)/(b - d)$.

Hence, if $H(x, y, z) = H(z, y, x)$, and $c_0 \neq c_2$, then

$$H(x, y, z) = \frac{(c_0 - c_2)[\Psi(x) - \Psi(z)]}{(c_0 - c_2)[\Phi(x) - \Phi(z)]}, \tag{26}$$

which is not symmetrical in the three variables. Then H is not symmetrical in x, y and z, unless, possibly, when $c_0 = c_2$.

Likewise from $H(x, y, z) = H(x, z, y)$, we are led to the conclusion that H is *not* a symmetrical function of x, y, and z, unless possibly when $c_1 = c_2$. But $c_0 = c_1 = c_2$ substituted into (15) makes $k_1 = k_2 = 0$, which is contrary to hypothesis that $k_r > 0$. Then in (25) the constants c_0, c_1 and c_2 can not be chosen in conformity with (15) so as to make $H(x, y, z)$ a symmetrical function of the *three* variables.

Symmetry in *two* variables will appear, however, if the mean (13) *reduces* to a mean of just *two* variables as it does when each $k_r = k$, constant, in which case,

$$M = \frac{\Psi(x_n) - \Psi(x_0)}{\Phi(x_n) - \Phi(x_0)}. \tag{27}$$

Although in the generalization (13) *symmetry* is thus lost, another property, *homogeneity* is retained in what seem to be the most important cases.

Most means $\Omega(x, y, \cdots, w)$ in common use are *homogeneous* functions of their arguments. That is, if c is a constant, and $\Omega(x, y, \cdots, w)$ and $\Omega(cx, cy, \cdots, cw)$ are both defined when $x, y, \cdots, w$ lie in some interval J, then

$$\Omega(cx, cy, \cdots, cw) = c\Omega(x, y, \cdots, w). \tag{28}$$

This *homogeneity* is associated geometrically with ruled surfaces, in particular with *cones*.

With reference to (8) and (10), let us write

$$F(x, y) = \frac{\Psi(y) - \Psi(x)}{\Phi(y) - \Phi(x)} \tag{29}$$

And now, let us consider a special variety of means obtained by taking in (8)

$$\phi(t) = t^q, \tag{30}$$

where q is any real number Then $F(x, y)$ is a homogeneous mean; that is,

$$F(cx, cy) = cF(x, y). \tag{31}$$

This is valid, indeed, even in the special cases, $q = 0$, -1, and -2, which lead, respectively to the arithmetic mean, the logarithmic mean (1) and to a second variety of logarithmic mean

$$m = \frac{xy \log (y/x)}{y - x}, \tag{32}$$

exhibited by Cisbani It may be noted that $q = -3/2$ leads to the geometric mean, and $q = -3$ to the harmonic mean of x and y

It is conceivable that for $\phi(t)$ other functions than t^q—functions not equivalent to t^q in integration—might be used to lead to a homogeneous $F(x, y)$ in (29). But such functions, if any, would hardly seem to be in common use.

The M in (13) retains the property of homogeneity, at least for $\phi(t) = t^q$; and so will also the more general means exhibited in the next section.

3. Further generalization. The means of Cisbani (5) suggest the following generalization. Let p be an integer or the reciprocal of an odd integer. With the notation of (13), take $k_r > 0$, and

$$F_p = \sum_1^n k_r f_r^p, \qquad G_p = \sum_1^n k_r g_r^p, \tag{33}$$

$$M_p = [G_p/F_p]^{1/p}. \tag{34}$$

Indeed, if in (8) and (10), $a \geqq 0$, then $g_r > 0$, and we may take for p any real number except zero. Now, M_p^p may be described as the weighted arithmetic mean of $(g_r/f_r)^p$ with *positive* weights $k_r f_r^p$. And hence M_p is an *internal* mean of x_0 , x_1 , $\cdots$, x_n ; that is

$$x_0 \leqq M_p \leqq x_n . \tag{35}$$

Furthermore, if in (8), $\phi(t) = t^q$, where q is any real number, then M_p is a homogeneous mean of x_0 , x_1 , $\cdots$, x_n .

Another generalization may be obtained by writing

$$m_r = g_r/f_r , \tag{36}$$

$$M'_p = [\Sigma k_r m_r^p / \Sigma k_r]^{1/p}. \tag{37}$$

And still another

$$M'' = [m_1^{k_1} \cdot m_2^{k_2} \cdots m_n^{k_n}]^{1/\Sigma k_r}. \tag{38}$$

These means (37) and (38) are internal; and they are homogeneous, if $F(x, y)$ in (29) is homogeneous.

The foregoing means are not, for $n > 1$, symmetrical functions of $x_1, x_2, \cdots, x_n$. Now the mere abandonment of (6) may lead to functions like (20) which are not means at all. But symmetry may be introduced as follows. First, lay aside (6), but suppose that the x_r are all different. Then let

$$f_{r,s} = \int_{x_r}^{x_s} \phi(t)\, dt, \qquad g_{r,s} = \int_{x_r}^{x_s} t\phi(t)\, dt; \tag{39}$$

where $r = 0, 1, \cdots, (n-1)$; $r < s \leqq n$. Then, let

$$U = \Sigma f_{r,s}^2 , \qquad V = \Sigma g_{r,s}^2 ; \tag{40}$$

where U and V is each a sum of $n(n-1)/2$ terms: Let W be the double-valued mean

$$W = \pm[V/U]^{1/2}. \tag{41}$$

Then W is a symmetric function of $x_0, x_1, \cdots, x_n$. If, in (8), $a' \leqq 0$, then in (12) each $g_r/f_r < 0$; and in (41) the negative value of W is an internal mean. But the positive radical is *external*. On the other hand, if $a \geqq 0$; then $g_r/f_r > 0$; and the positive radical in (41) is internal. In this case, it may be well to use for W *only the positive value of* W.

In the more general case where $a < 0$ and $a' > 0$, the fractions g_r/f_r may have different signs. But, in all cases, *at least one* of the two radicals (41) is an *internal* mean of $x_0, x_1, \cdots, x_n$. Moreover, W is homogeneous, if in (8), $\phi(t) = t^q$.

Finally, let

$$m_{r,s} = g_{r,s}/f_{r,s} , \tag{42}$$

$$Z = \pm\{[\Sigma m_{r,s}^2]/n(n-1)\}^{1/2}. \tag{43}$$

Then Z is symmetric; and at least one value is internal. If $a > 0$, we would naturally take $Z > 0$; and this Z is then an internal mean. Moreover, Z is homogeneous if the $m_{r,s}$ are homogeneous; that is, if $F(x, y)$ in (29) is homogeneous for every x and y in I.

A STUDY OF R. A. FISHER'S z DISTRIBUTION AND THE RELATED F DISTRIBUTION[1]

By Leo A. Aroian

Hunter College

1. Nature of the problem. Consider two samples of N_1 and N_2 drawings, each sample drawn from one of two populations consisting of variates normally distributed with equal population variances σ^2. We define the two sample means $\bar{x}_1 = \dfrac{\sum_{i=1}^{N_1} x_i}{N_1}$, $\bar{x}_2 = \dfrac{\sum_{j=1}^{N_2} x_j}{N_2}$, x_i's and x_j's independent variates. We calculate from the two samples

$$s_1^2 = \frac{\sum_{i=1}^{N_1} (x_i - \bar{x}_1)^2}{n_1} \quad \text{and} \quad s_2^2 = \frac{\sum_{j=1}^{N_2} (x_j - \bar{x}_2)^2}{n_2}, \qquad n_1 = N_1 - 1,\ n_2 = N_2 - 1.$$

The distribution of $z = \frac{1}{2} \log \dfrac{s_1^2}{s_2^2}$ is well known.

$$P(z) = \frac{2 n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2}}{\mathrm{B}\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \frac{e^{n_1 z}}{(n_1 e^{2z} + n_2)^{\frac{1}{2}(n_1+n_2)}} dz. \tag{1.1}$$

We shall denote the ordinates by $y(z)$. The purpose of this study is to discuss the seminvariants of the z distribution and also to find useful approximations for them; to show that as n_1 and n_2 approach infinity in any manner whatever the distribution of z approaches normality; to find the upper bound of the absolute value of the difference between the distribution function of z and the function determined by the approximate seminvariants of the distribution of z for n_1 and n_2 large; to approximate the z distribution by the Type III distribution, the Gram-Charlier Type A series, and the logarithmic frequency curve; and finally to investigate the same properties with respect to the F distribution, where $F = e^{2z} = \dfrac{s_1^2}{s_2^2}$. The non-existence of the moments of F for certain values of n_1 and n_2 is noted and explained on the basis of the distribution of the quotient $\dfrac{y}{x}$.

---

[1] Presented to the American Mathematical Society, September 10, 1938, New York City in part; and to the Institute December 27, 1939 at Philadelphia.

2. General features of the z distribution. The z distribution is always unimodal, asymmetrical if $n_1 \neq n_2$, and symmetrical if $n_1 = n_2$. We see that interchanging n_1 and n_2 is the same as replacing z by $-z$. Fisher [7] noted that the two parameter family of curves includes as special cases the normal curve, the χ^2 distribution, and Student's distribution. The mode is at $z = 0$, the maximum ordinate is

$$y(0) = \frac{2n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} (n_1 + n_2)^{-\frac{1}{2}(n_1+n_2)}$$

or approximately

$$y(0) = \frac{1}{\sqrt{2\pi}} \left\{\frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}^{\frac{1}{2}} \quad \text{for } n_1 \text{ and } n_2 \text{ large.} \tag{2.1}$$

The two points of inflection are

$$z = \tfrac{1}{2} \log \left\{\frac{n_1 n_2 + n_1 + n_2 \pm \sqrt{n_1^2 + n_2^2 + 2n_1^2 n_2 + 2n_1 n_2^2 + 2n_1 n_2}}{n_1 n_2}\right\}. \tag{2.2}$$

They are equidistant from the mode, a property also of the Pearson system of frequency curves [24]. Also $\lim_{z \to \pm\infty} z^n \frac{d^n y(z)}{dz^n} = 0$.

3. The moment generating function and seminvariants. The moment generating function of the z distribution is

$$M_z(\theta) = \left(\frac{n_2}{n_1}\right)^{\frac{1}{2}\theta} \frac{B\left(\frac{n_2 - \theta}{2}, \frac{n_1 + \theta}{2}\right)}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} = \left(\frac{n_2}{n_1}\right)^{\frac{1}{2}\theta} \frac{\Gamma\left(\frac{n_2 - \theta}{2}\right)\Gamma\left(\frac{n_1 + \theta}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}. \tag{3.1}$$

The seminvariants of Thiele are defined by the following identity in θ:

$$\log M_z(\theta) = \lambda_1 \theta + \lambda_2 \frac{\theta^2}{2!} + \lambda_3 \frac{\theta^3}{3!} + \lambda_4 \frac{\theta^4}{4!} + \cdots. \tag{3.2}$$

To find λ_r we take the logarithm of the moment generating function, expand it in powers of θ and choose the coefficient of $\frac{\theta^r}{r!}$. A complete discussion of properties of seminvariants may be found elsewhere [4].

4. The seminvariants of z. Now by the following formulas [11] p. 38:

$$\log \Gamma(1 + x) = \frac{-s_1 x}{1} + \frac{s_2 x^2}{2} - \frac{s_3 x^3}{3} + \frac{s_4 x^4}{4} - \cdots, \qquad |x| < 1, \tag{4.1}$$

$$(4\,2)\quad \log \Gamma(1-x) = s_1 x + \frac{s_2 x^2}{2} + \frac{s_3 x^3}{3} + \frac{s_4 x^4}{4} + \cdots, \qquad |x| < 1,$$

where in both formulas

$$s_1 = \lim_{n\to\infty}\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} - \log n\right),$$

$$s_n = \frac{1}{1^n} + \frac{1}{2^n} + \frac{1}{3^n} + \frac{1}{4^n} + \cdots, \qquad n \geqq 2.$$

Also

$$(4.3)\quad \log \mathrm{B}(\tfrac{1}{2}[1+x], \tfrac{1}{2}) = \log \pi - \sigma_1 x + \sigma_2 \frac{x^2}{2} - \sigma_3 \frac{x^3}{3} + \sigma_4 \frac{x^4}{4} - \cdots, \qquad |x| < 1,$$

where

$$\sigma_n = \frac{1}{1^n} - \frac{1}{2^n} + \frac{1}{3^n} - \frac{1}{4^n} + \cdots, \qquad n \geqq 1$$

and

$$\sigma_n = \left(1 - \frac{1}{2^{n-1}}\right) s_n, \qquad n \geqq 2.$$

Hence from (4 1) and (4 3)

$$(4.4)\quad \log \Gamma\left(\frac{1+x}{2}\right) = \tfrac{1}{2}\log \pi - x\left(\sigma_1 + \frac{s_1}{2}\right) + \frac{x^2}{2}\left(\sigma_2 + \frac{s_2}{2^2}\right) - \frac{x^3}{3}\left(\sigma_3 + \frac{s_3}{2^3}\right) + \frac{x^4}{4}\left(\sigma_4 + \frac{s_4}{2^4}\right) - \cdots.$$

Since $\sigma_n = \left(1 - \frac{1}{2^{n-1}}\right) s_n$, $n \geqq 2$, we may write (4.4) as

$$(4\,5)\quad \log \Gamma\left(\frac{1+x}{2}\right) = \tfrac{1}{2}\log \pi - x\left(\sigma_1 + \frac{s_1}{2}\right) + \sum_{k=2}^{\infty} \frac{(-1)^k x^k}{k}\left(1 - \frac{1}{2^k}\right) s_k.$$

From (3.1)

$$(4.6)\quad \log M_z(\theta) = \log \Gamma\left(\frac{n_2 - \theta}{2}\right) + \log \Gamma\left(\frac{n_1 + \theta}{2}\right) + \frac{\theta}{2}(\log n_2 - \log n_1) - \log \Gamma\left(\frac{n_1}{2}\right) - \log \Gamma\left(\frac{n_2}{2}\right).$$

The results assume slightly different forms for (A) n_1 and n_2 each even; (B) n_1 and n_2 each odd; (C) n_1 even, n_2 odd, (D) n_1 odd, n_2 even The general formula for $\lambda_{,z}$ for all cases is

$$\lambda_{r:s} = \sum_{k=0}^{\infty} \left\{ \frac{(-1)^r (r-1)!}{(n_1+2k)^r} + \frac{(r-1)!}{(n_2+2k)^r} \right\}, \qquad r \geqq 2. \tag{4.7}$$

This result is not so useful from the point of view of numerical applications as the formulas which follow.

5. Case A, n_1 and n_2 each even. From (4.6)

$$\log \Gamma\left(\frac{n_2-\theta}{2}\right) = \log\left(\frac{n_2-2-\theta}{2}\right) + \log\left(\frac{n_2-4-\theta}{2}\right) + \cdots + \log\left(1-\frac{\theta}{2}\right) + \log \Gamma\left(1-\frac{\theta}{2}\right). \tag{5.1}$$

Now $\log\left(1-\frac{\theta}{n_2-2}\right) = -\sum_{k=1}^{\infty} \frac{1}{k}\left(\frac{\theta}{n_2-2}\right)^k$. There will be $\frac{n_2}{2}-1$ series of this sort, and only one series of the type $\log \Gamma\left(1-\frac{\theta}{2}\right) = \sum_{k=1}^{\infty} \frac{s_k}{k}\left(\frac{\theta}{2}\right)^k$ as given by (4.1). In the above expansion and those succeeding, terms not involving θ are omitted, since such terms are not needed in finding the seminvariants of z. The series $\log \Gamma\left(1-\frac{\theta}{2}\right)$ will always occur. Then

$$\log \Gamma\left(\frac{n_2-\theta}{2}\right) = -\sum_{k=1}^{\infty} \frac{1}{k}\left[\left(\frac{\theta}{n_2-2}\right)^k + \left(\frac{\theta}{n_2-4}\right)^k + \cdots + \left(\frac{\theta}{2}\right)^k - s_k\left(\frac{\theta}{2}\right)^k\right], \tag{5.2}$$

or

$$\log \Gamma\left(\frac{n_2-\theta}{2}\right) = \sum_{k=1}^{\infty} \frac{s_k}{k}\left(\frac{\theta}{2}\right)^k - \sum_{k=1}^{\infty} \frac{1}{k} \sum_{l=1}^{\frac{1}{2}n_2-1} \left(\frac{\theta}{2l}\right)^k. \tag{5.3}$$

We remark that the double sum is zero if $n_2 = 2$. Similarly

$$\log \Gamma\left(\frac{n_1+\theta}{2}\right) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}\left\{\left(\frac{\theta}{n_1-2}\right)^k + \left(\frac{\theta}{n_1-4}\right)^k + \cdots + \left(\frac{\theta}{2}\right)^k - s_k\left(\frac{\theta}{2}\right)^k\right\}, \tag{5.4}$$

or

$$\log \Gamma\left(\frac{n_1+\theta}{2}\right) = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} s_k\left(\frac{\theta}{2}\right)^k + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \sum_{l=1}^{\frac{1}{2}n_1-1} \left(\frac{\theta}{2l}\right)^k. \tag{5.5}$$

By use of (5.3) and (5.5) we have for the seminvariants of z, when n_1 and n_2 are even

$$\lambda_{r:s} = \frac{(r-1)!}{2^r}\left\{\left(s_r - \sum_{k=1}^{\frac{1}{2}n_2-1} \frac{1}{k^r}\right) + (-1)^r\left(s_r - \sum_{k=1}^{\frac{1}{2}n_1-1} \frac{1}{k^r}\right)\right\}, \qquad r \geqq 2. \tag{5.6}$$

For $\lambda_{1\cdot z} = \bar{z}$ we have by (4.6), (4.3), and (4.5)

$$\lambda_{1.z} = \frac{1}{2}\left[\left(\log n_2 - \sum_{k=1}^{\frac{1}{2}n_2-1} \frac{1}{k}\right) - \left(\log n_1 - \sum_{k=1}^{\frac{1}{2}n_1-1} \frac{1}{k}\right)\right]. \tag{5.7}$$

6. Case B, n_1 and n_2 odd. We have

$$\log \Gamma\left(\frac{n_2-\theta}{2}\right) = \log\left(\frac{n_2-2-\theta}{2}\right) + \log\left(\frac{n_2-4-\theta}{2}\right) + \cdots + \log\left(\frac{1-\theta}{2}\right) + \log \Gamma\left(\frac{1-\theta}{2}\right). \tag{6.1}$$

Expanding $\log \Gamma\left(\frac{1-\theta}{2}\right)$ by (4.5)

$$\log \Gamma\left(\frac{n_2-\theta}{2}\right) = -\left[\sum_{k=1}^{\infty} \frac{\theta^k}{k(n_2-2)^k} + \frac{\theta^k}{k(n_2-4)^k} + \cdots + \frac{\theta^k}{k}\right] + \theta\left(\sigma_1 + \frac{s_1}{2}\right) + \sum_{k=2}^{\infty} \frac{\theta^k}{k}\left(1 - \frac{1}{2^k}\right) s_k. \tag{6.2}$$

However $s_k\left(1 - \frac{1}{2^k}\right) = \frac{1}{1^k} + \frac{1}{3^k} + \frac{1}{5^k} + \frac{1}{7^k} + \cdots$, $k > 1$, which we shall denote hereafter by t_k. Hence (6.2) becomes

$$\log \Gamma\left(\frac{n_2-\theta}{2}\right) = \theta\left(\sigma_1 + \frac{s_1}{2}\right) + \sum_{k=2}^{\infty} \frac{\theta^k}{k} t_k - \sum_{k=1}^{\infty} \frac{1}{k} \sum_{l=0}^{\frac{1}{2}(n_2-3)} \left(\frac{\theta}{2l+1}\right)^k. \tag{6.3}$$

Also

$$\log \Gamma\left(\frac{n_1+\theta}{2}\right) = \log\left(\frac{n_1+\theta-2}{2}\right) + \log\left(\frac{n_1+\theta-4}{2}\right) + \cdots + \log\left(\frac{1+\theta}{2}\right) + \log \Gamma\left(\frac{1+\theta}{2}\right), \tag{6.4}$$

and

$$\log \Gamma\left(\frac{n_1+\theta}{2}\right) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}\left[\frac{\theta^k}{(n_1-2)^k} + \frac{\theta^k}{(n_1-4)^k} + \cdots + \frac{\theta^k}{1}\right] - \theta\left(\sigma_1 + \frac{s_1}{2}\right) + \sum_{k=2}^{\infty} \frac{(-1)^k}{k} \theta^k t_k. \tag{6.5}$$

$$\log \Gamma\left(\frac{n_1+\theta}{2}\right) = -\theta\left(\sigma_1 + \frac{s_1}{2}\right) + \sum_{k=2}^{\infty} \frac{(-1)^k \theta^k}{k} t_k + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \sum_{l=0}^{\frac{1}{2}(n_1-3)} \frac{\theta^k}{(2l+1)^k}. \tag{6.6}$$

Combining both these results (6.3) and (6 6) we have

$$\lambda_{r,z} = (r-1)!\left\{\left(t_r - \sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(2k+1)^r}\right) + (-1)^r\left(t_r - \sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{(2k+1)^r}\right)\right\}, \qquad r \geqq 2. \tag{6.7}$$

$$\lambda_{1,z} = \bar{z} = \left(\frac{1}{2}\log n_2 - \sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{2k+1}\right) - \left(\frac{1}{2}\log n_1 - \sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{2k+1}\right) \tag{6.8}$$

7. Cases C, D, and values of s_k, σ_1, t_k. The formulas for case C, n_1 even, n_2 odd are

$$\lambda_{r,z} = (r-1)!\left\{\left(t_r - \sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(2k+1)^r}\right) + \frac{(-1)^r}{2^r}\left(s_r - \sum_{k=1}^{\frac{1}{2}n_1-1} \frac{1}{k^r}\right)\right\}, \qquad r \geqq 2. \tag{7.1}$$

$$\lambda_{1,z} = \bar{z} = \frac{1}{2}\log\frac{n_2}{n_1} + \frac{1}{2}\sum_{k=1}^{\frac{1}{2}n_1-1}\frac{1}{k} - \sum_{k=0}^{\frac{1}{2}(n_2-3)}\frac{1}{2k+1} + \sigma_1. \tag{7.2}$$

The results for case D, n_1 odd, n_2 even are

$$\lambda_{r,z} = (r-1)!\left\{\frac{1}{2^r}\left(s_r - \sum_{k=1}^{\frac{1}{2}n_2-1}\frac{1}{k^r}\right) + (-1)^r\left(t_r - \sum_{k=0}^{\frac{1}{2}(n_1-3)}\frac{1}{(2k+1)^r}\right)\right\}, \qquad r \geqq 2. \tag{7.3}$$

$$\lambda_{1,z} = \bar{z} = \frac{1}{2}\log\frac{n_2}{n_1} - \sigma_1 + \sum_{k=0}^{\frac{1}{2}(n_1-3)}\frac{1}{2k+1} - \frac{1}{2}\sum_{k=1}^{\frac{1}{2}n_2-1}\frac{1}{k}. \tag{7.4}$$

We list the numerical values of s_k and t_k, $k \leqq 10$. The values of s_k are from Stieltjes [20],

$$\begin{aligned} s_1 &= 0.57721\ 56649 \\ s_2 &= 1.64493\ 40668 \\ s_3 &= 1.20205\ 69032 \\ s_4 &= 1.08232\ 32337 \\ s_5 &= 1.03692\ 77551 \\ s_6 &= 1.01734\ 30620 \\ s_7 &= 1.00834\ 92774 \\ s_8 &= 1.00407\ 73562 \\ s_9 &= 1.00200\ 83928 \\ s_{10} &= 1.00099\ 45751 \end{aligned} \tag{7.5}$$

$$\begin{aligned} \sigma_1 &= \log 2 = 0.69317\ 0206 \\ t_2 &= 1.23370\ 00550 \\ t_3 &= 1.05179\ 97903 \\ t_4 &= 1.01467\ 80316 \\ t_5 &= 1.00452\ 37628 \\ t_6 &= 1.00144\ 70767 \\ t_7 &= 1.00047\ 15487 \\ t_8 &= 1.00015\ 51790 \\ t_9 &= 1.00005\ 13452 \\ t_{10} &= 1.00001\ 70113 \end{aligned} \tag{7.6}$$

By means of the formula $t_k = s_k\left(1 - \frac{1}{2^k}\right)$, $k > 1$, t_k was calculated from s_k. From the well known results for the Zeta function of Riemann $\zeta(s)$, [22], (p. 265, p. 267),

$$\zeta_s = s_k = \sum_{k=1}^{\infty}\frac{1}{k^s} = \frac{1}{\Gamma(s)}\int_0^{\infty}\frac{x^{s-1}e^{-x}}{1-e^{-x}}\,dx, \qquad s \geqq 1, \qquad k > 1. \tag{7.7}$$

$$\sigma_s = \left(1 - \frac{1}{2^{s-1}}\right)\zeta(s) = \frac{1}{\Gamma(s)}\int_0^\infty \frac{x^{s-1}}{e^x + 1}\,dx, \qquad \text{and} \tag{7.8}$$

$$t_s = \zeta(s)\left(1 - \frac{1}{2^s}\right). \tag{7.9}$$

8. The mean of the z distribution. From our previous formulas for $\bar{z}$ we prove that if $n_1 = n_2$, $\bar{z} = 0$, and $\bar{z} < 0$ for $n_2 > n_1$, $\bar{z} > 0$ for $n_1 > n_2$ The maximum absolute value of $\lambda_{1,z}$ will occur when $n_1 = 1$, $n_2 = \infty$, or $n_1 = \infty$, $n_2 = 1$, and from (7.4) or (6 8) we have $\max |\lambda_{1,z}| = \frac{s_1}{2} + \frac{1}{2}\log 2 = .6352$

9. Formulas for $\lambda_{2,z}$, $\mu_{2,z}$, $\lambda_{3,z}$, $\mu_{3,z}$, $\lambda_{4,z}$, and $\mu_{4,z}$. We have four cases from (5.6), (6 7), (7 1), (7 3):

$$\lambda_{2,z} = \frac{1}{4}\left[2s_2 - \sum_{k=1}^{\frac{1}{2}(n_1-2)}\frac{1}{k^2} - \sum_{k=1}^{\frac{1}{2}(n_2-2)}\frac{1}{k^2}\right] \tag{9.1}$$

$$= .822467 - \frac{1}{4}\left(\sum_{k=1}^{\frac{1}{2}(n_1-2)}\frac{1}{k^2} + \sum_{k=1}^{\frac{1}{2}(n_2-2)}\frac{1}{k^2}\right), \qquad n_1, n_2 \text{ even.}$$

$$\lambda_{2,z} = 2.467401 - \frac{1}{4}\left(\sum_{k=0}^{\frac{1}{2}(n_2-3)}\frac{1}{(k+\frac{1}{2})^2} + \sum_{k=0}^{\frac{1}{2}(n_1-3)}\frac{1}{(k+\frac{1}{2})^2}\right), \qquad n_1, n_2 \text{ odd.} \tag{9.2}$$

$$\lambda_{2,z} = 1\ 644934 - \frac{1}{4}\left(\sum_{k=1}^{\frac{1}{2}(n_1-1)}\frac{1}{k^2} + \sum_{k=0}^{\frac{1}{2}(n_2-3)}\frac{1}{(k+\frac{1}{2})^2}\right), \qquad n_1 \text{ even, } n_2 \text{ odd.} \tag{9.3}$$

$$\lambda_{2,z} = 1\ 644934 - \frac{1}{4}\left(\sum_{k=0}^{\frac{1}{2}(n_1-3)}\frac{1}{(k+\frac{1}{2})^2} + \sum_{k=1}^{\frac{1}{2}(n_2-2)}\frac{1}{k^2}\right), \qquad n_1 \text{ odd, } n_2 \text{ even.} \tag{9.4}$$

In all cases of course $\lambda_{2,z} > 0$ and moreover $\lambda_{2,z} \to 0$ as n_1 and $n_2 \to \infty$. We list

$$\lambda_{3,z} = \frac{1}{4}\left(\sum_{k=1}^{\frac{1}{2}n_1-1}\frac{1}{k^3} - \sum_{k=1}^{\frac{1}{2}n_2-1}\frac{1}{k^3}\right), \qquad n_1, n_2 \text{ even.} \tag{9.5}$$

$$\lambda_{3,z} = \frac{1}{4}\left(\sum_{k=0}^{\frac{1}{2}(n_1-3)}\frac{1}{(k+\frac{1}{2})^3} - \sum_{k=0}^{\frac{1}{2}(n_2-3)}\frac{1}{(k+\frac{1}{2})^3}\right), \qquad n_1, n_2 \text{ odd.} \tag{9.6}$$

$$\lambda_{3,z} = 1\ 803085 + \frac{1}{4}\left(\sum_{k=1}^{\frac{1}{2}(n_1-2)}\frac{1}{k^3} - \sum_{k=0}^{\frac{1}{2}(n_2-3)}\frac{1}{(k+\frac{1}{2})^3}\right), \qquad n_1 \text{ even, } n_2 \text{ odd} \tag{9.7}$$

$$\lambda_{3,z} = -1.803085 + \frac{1}{4}\left(\sum_{k=0}^{\frac{1}{2}(n_1-3)}\frac{1}{(k+\frac{1}{2})^3} - \sum_{k=1}^{\frac{1}{2}(n_2-2)}\frac{1}{k^3}\right), \qquad n_1 \text{ odd, } n_2 \text{ even.} \tag{9 8}$$

$$\lambda_{4,z} = 811742 - \frac{3}{8}\left(\sum_{k=1}^{\frac{1}{2}n_2-1}\frac{1}{k^4} + \sum_{k=1}^{\frac{1}{2}n_1-1}\frac{1}{k^4}\right), \qquad n_1, n_2 \text{ even.} \tag{9.9}$$

$$\lambda_{4,z} = 12\ 17614 - 6\left(\sum_{k=0}^{\frac{1}{2}(n_2-3)}\frac{1}{(2k+1)^4} + \sum_{k=0}^{\frac{1}{2}(n_1-3)}\frac{1}{(2k+1)^4}\right), \qquad n_1, n_2 \text{ odd} \tag{9 10}$$

$$\lambda_{4,s} = 6.493939 - 6\left(\sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(2k+1)^4} + \sum_{k=1}^{\frac{1}{2}(n_1-1)} \frac{1}{k^4}\right), \qquad n_1 \text{ even}, n_2 \text{ odd} \tag{9.11}$$

$$\lambda_{4,s} = 6.493939 - 6\left(\sum_{k=1}^{\frac{1}{2}(n_2-2)} \frac{1}{k^4} + \sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{(2k+1)^4}\right), \qquad n_1 \text{ odd}, n_2 \text{ even}. \tag{9.12}$$

We see $\lambda_{r,s} > 0$ whenever r is even. If r is odd $\lambda_{r,s} < 0$ if $n_2 > n_1$, and $\lambda_{r,s} > 0$ if $n_1 > n_2$. Also $\mu_{r,s} > 0$, $n_1 > n_2$, r odd, greater than one. Similarly $\mu_{r,s} < 0$, r odd > 1, $n_2 > n_1$.

10. Skewness, excess, and values of α_n. We take for our measure of skewness $\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\lambda_3}{\lambda_2^{3/2}}$. For $n_2 > n_1$, $\alpha_3 < 0$. Further the skewness increases negatively if n_1 remains constant as $n_2 \to \infty$. Thus negative skewness will be a maximum for $n_2 = \infty$, $n_1 = 1$, and positive skewness will be a maximum when $n_2 = 1$, $n_1 = \infty$. The absolute value of maximum α_3 is

$$|\alpha_3| = \left|\frac{2l_3}{l_1^{3/2}}\right| = 1.5351. \tag{10.1}$$

As our measure of kurtosis we use $\alpha_4 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{\lambda_4}{\lambda_2^2}$. As a measure of excess, E, we use $E = \alpha_4 - 3 = \frac{\lambda_4}{\lambda_2^2}$. The excess is always positive.

11. Approximations for $\lambda_{r,s}$ by the Euler-Maclaurin sum formula. The exact results given previously for the seminvariants become unwieldy for n_1 and n_2 large. Hence we develop useful approximations for the seminvariants, and give the maximum error of the approximation. We find first our results for $\lambda_{r,s}$ when n_1 and n_2 are even and $r > 1$. We begin with (5.6)

$$\lambda_{r,s} = \frac{(r-1)!}{2^r}\left\{\left(s_r - \sum_{k=1}^{\frac{1}{2}n_2-1} \frac{1}{k^r}\right) + (-1)^r\left(s_r - \sum_{k=1}^{\frac{1}{2}n_1-1} \frac{1}{k^r}\right)\right\}$$

and rewrite this as

$$\lambda_{r,s} = \frac{(r-1)!}{2^r}\left\{\sum_{k=\frac{1}{2}n_2}^{\infty} \frac{1}{k^r} + (-1)^r \sum_{k=\frac{1}{2}n_1}^{\infty} \frac{1}{k^r}\right\}. \tag{11.1}$$

Now find the two sums of (11.1) by the Euler-Maclaurin sum formula [21] using the first three terms, and obtain

$$\begin{aligned}\lambda_{r,s} = \frac{(r-2)!}{2}\Bigg[&\left(\frac{n_2+r-1}{n_2^r} + (-1)^r\frac{n_1+r-1}{n_1^r}\right) \\ &+ \frac{r(r-1)}{3}\left(\frac{1}{n_2^{r+1}} + \frac{(-1)^r}{n_1^{r+1}}\right) \\ &- \frac{r(r-1)(r+1)(r+2)}{45}\left(\frac{1}{n_2^{r+3}} + \frac{(-1)^r}{n_1^{r+3}}\right)\Bigg].\end{aligned} \tag{11.2}$$

We use the following theorem [10] (p. 539), to find the error:

If $f(x)$ is of constant sign for $x > 0$, and together with all of its derivatives, tends monotonely to zero as $x \to \infty$, Euler's summation formula may be stated in the simplified form

$$\sum_{x=0}^{n} f_x = \int_0^n f(x)\,dx + \tfrac{1}{2}(f_n + f_0) + \frac{B_2}{2!}(f_n' - f_0') + \cdots + \frac{(-1)^{k-1}B_{2k}}{(2k)!}(f_n^{(2k-1)} - f_0^{(2k-1)}) + \frac{\theta B_{2k+2}}{(2k+2)!}(f_n^{(2k+1)} - f_0^{(2k+1)})$$

where $0 < \theta < 1$ and $B_2 = 1/6$, $B_4 = 1/30$, $B_6 = 1/42$, $B_8 = 1/30$, $B_{10} = 5/66$, etc. If we use

$$\lambda_{r,z} = \frac{(r-2)!}{2}\left(\frac{n_2 + r - 1}{n_2^r} + (-1)^r \frac{n_1 + r - 1}{n_1^r}\right), \tag{11.3}$$

then the error committed is of the same sign and less than

$$\frac{r!}{3!}\left\{\frac{1}{n_2^{r+1}} + \frac{(-1)^r}{n_1^{r+1}}\right\}.$$

If we take

$$\lambda_{r,z} = \frac{(r-2)!}{2}\left[\left(\frac{n_2 + r - 1}{n_2^r} + (-1)^r \frac{n_1 + r - 1}{n_1^r}\right) + \frac{r(r-1)}{3}\left(\frac{1}{n_2^{r+1}} + \frac{(-1)^r}{n_1^{r+1}}\right)\right], \tag{11.4}$$

then our error is less than, and has the same sign as

$$-\frac{(r+2)!}{90}\left\{\frac{1}{n_2^{r+3}} + \frac{(-1)^r}{n_1^{r+3}}\right\}$$

Finally if we use (11.2), our error has the same sign as, and is less than

$$\frac{(r+4)!}{945}\left\{\frac{1}{n_2^{r+5}} + \frac{(-1)^r}{n_1^{r+5}}\right\}$$

12. Approximations for other values of n_1 and n_2, $r > 1$. Now in case n_1 and n_2 are odd we have from (6.7)

$$\lambda_{r,z} = (r-1)!\left\{\sum_{k=\frac{1}{2}(n_2-1)}^{\infty} \frac{1}{(2k+1)^r} + (-1)^r \sum_{k=\frac{1}{2}(n_1-1)}^{\infty} \frac{1}{(2k+1)^r}\right\}. \tag{12.1}$$

Applying the Euler-Maclaurin sum formula to each of the sums in (12.1) we are led to exactly the same results given in paragraph (11). The other cases are obvious combinations of the sums in (11.1) and (12.1), and so for all values of n_1 and n_2 the approximate results for $\lambda_{r,z}$, $r > 1$ are

$$\lambda_{r,s} = \frac{(r-2)!}{2}\left\{\frac{n_2+r-1}{n_2^r} + (-1)^r \frac{n_1+r-1}{n_1^r}\right\} + \frac{r!}{6}\left\{\frac{1}{n_2^{r+1}} + \frac{(-1)^r}{n_1^{r+1}}\right\} - \frac{(r+2)!}{90}\left\{\frac{1}{n_2^{r+3}} + \frac{(-1)^r}{n_1^{r+3}}\right\}. \tag{12.2}$$

Formulas (11.1) and (12.1) prove the result previously given for $\lambda_{r,s}$ (4.7).

13. The approximate values of $\lambda_{1,s}$. From (5.7)

$$\lambda_{1,s} = \frac{1}{2}\left[\left(\log n_2 - \sum_{k=1}^{\frac{1}{2}n_2-1}\frac{1}{k}\right) - \left(\log n_1 - \sum_{k=1}^{\frac{1}{2}n_1-1}\frac{1}{k}\right)\right], \qquad n_1 \text{ and } n_2 \text{ even.}$$

We use the Euler-Maclaurin sum formula on the sum

$$\sum_{k=1}^{\frac{1}{2}n_2-1}\frac{1}{k} = \left\{\sum_{k=0}^{\frac{1}{2}n_2-1}\left(\frac{1}{k+1}\right) - \frac{2}{n_2}\right\}$$

and the similar sum involved in $\lambda_{1,s}$. Hence we have

$$\lambda_{1,s} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) + \frac{1}{6}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right) - \frac{1}{15}\left(\frac{1}{n_2^4} - \frac{1}{n_1^4}\right), \qquad n_1, n_2 > 2. \tag{13.1}$$

The errors committed by using one, two, or three terms of (13.1) are less than, and of the same sign respectively as

$$\frac{1}{6}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right), \qquad -\frac{1}{15}\left(\frac{1}{n_2^4} - \frac{1}{n_1^4}\right), \qquad \frac{8}{63}\left(\frac{1}{n_2^6} - \frac{1}{n_1^6}\right).$$

For n_1 and n_2 both odd we find the same result as (13.1). The restriction $n_1, n_2 > 2$, may easily be replaced by $n_1, n_2 \geqq 2$ (for n_1, n_2 even) and $n_1, n_2 \geqq 1$ (for n_1, n_2 both odd). When n_1 is odd, n_2 even, the formula is again the same as (13.1) if n_1 and n_2 are sufficiently large; but if n_1 and n_2 are small we find in this case

$$\lambda_{1,s} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) + \frac{1}{6}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right) - \frac{1}{15}\left(\frac{1}{n_2^4} - \frac{1}{n_1^4}\right) + \frac{1}{2}\left(1 - \frac{1}{2}\right) + \frac{1}{6}\left(1 - \frac{1}{4}\right) - \frac{1}{15}\left(1 - \frac{1}{16}\right) - \frac{1}{2}\log 2. \tag{13.2}$$

Another method of finding (12.2) would have been to use the asymptotic expression for $\log \Gamma(x)$.

14. Approximate values of $\lambda_{r,s}$ for values of r. We list the approximate values of $\lambda_{r,s}$ to three terms.

$$\lambda_{1,s} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) + \frac{1}{6}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right) - \frac{1}{15}\left(\frac{1}{n_2^4} - \frac{1}{n_1^4}\right)$$

$$
\begin{aligned}
\lambda_{2\cdot z} &= \frac{1}{2}\left(\frac{n_2+1}{n_2^2}+\frac{n_1+1}{n_1^2}\right)+\frac{1}{3}\left(\frac{1}{n_2^3}+\frac{1}{n_1^3}\right)-\frac{4}{15}\left(\frac{1}{n_2^5}+\frac{1}{n_1^5}\right)\\
\lambda_{3\cdot z} &= \frac{1}{2}\left(\frac{n_2+2}{n_2^3}-\frac{n_1+2}{n_1^3}\right)+\left(\frac{1}{n_2^4}-\frac{1}{n_1^4}\right)-\frac{4}{3}\left(\frac{1}{n_2^6}-\frac{1}{n_1^6}\right)\\
\lambda_{4\cdot z} &= \left(\frac{n_2+3}{n_2^4}+\frac{n_1+3}{n_1^4}\right)+4\left(\frac{1}{n_2^6}+\frac{1}{n_1^5}\right)-8\left(\frac{1}{n_2^7}+\frac{1}{n_1^7}\right)\\
\lambda_{5\cdot z} &= 3\left(\frac{n_2+4}{n_2^5}-\frac{n_1+4}{n_1^5}\right)+20\left(\frac{1}{n_2^6}-\frac{1}{n_1^6}\right)-56\left(\frac{1}{n_2^8}-\frac{1}{n_1^8}\right)\\
\lambda_{6\cdot z} &= 12\left(\frac{n_2+5}{n_2^6}+\frac{n_1+5}{n_1^6}\right)+120\left(\frac{1}{n_2^7}+\frac{1}{n_1^7}\right)-448\left(\frac{1}{n_2^9}+\frac{1}{n_1^9}\right).
\end{aligned}
\tag{14.1}
$$

The approximate values given by Cornish and Fisher [8] (p. 319), are similar, but have fewer terms. Cornish and Fisher give no remainder term From (14 1) and (12 2) we see the maximum absolute values of $\lambda_{2r+1.z}$, $r \geqq 1$, occur when $n_2 = \infty$, $n_1 = 1$, or $n_2 = 1$, $n_1 = \infty$. Similarly $\lambda_{2r\,z}$, $r \geqq 1$, has its maximum value for $n_1 = n_2 = 1$ The standard seminvariants of z are defined $\xi_{r.z} = \dfrac{\lambda_r}{\lambda_2^{\frac{1}{2}r}}$, $r \geqq 2$. We also note that for $n_2 > n_1$, $\xi_{2r+1\,z} < 0$, $r \geqq 1$ and hence $\alpha_{2r+1} < 0$ also where $\alpha_n = \dfrac{\mu_n}{\mu_2^{\frac{1}{2}n}}$ Moreover the maximum absolute values of $\xi_{2r:z}$ and $\xi_{2r+1\,z}$ occur when $n_1 = 1$, $n_2 = \infty$ or $n_2 = 1$, $n_1 = \infty$, and also for α_{2r} and α_{2r+1}. Approximately then

$$
\max \xi_{r.z} = (-1)^r \frac{(r-1)!}{2}, \qquad r \geqq 2. \tag{14.2}
$$

The exact value for maximum $\alpha_{4.z}$ is $3 + \dfrac{6t_4}{t_2^2} = 7.07$.

15. Approach to normality of the z distribution. We prove the theorem: The distribution of z approaches normality as n_1 and $n_2 \to \infty$ in any manner whatever, with $\bar{z} = \dfrac{1}{2}\left(\dfrac{1}{n_2}-\dfrac{1}{n_1}\right)$, $\sigma_z^2 = \dfrac{1}{2}\left(\dfrac{1}{n_2}+\dfrac{1}{n_1}\right)$ We also find an upper bound of the absolute value of the difference between the z distribution and the function determined by the approximate seminvariants of z when n_1 and n_2 become large. To prove the theorem we start with the original distribution of z, and find when n_1 and n_2 are large,

$$
P(z) = \frac{1}{\sqrt{2\pi}\,\sigma_z}\left\{\frac{n_1+n_2}{n_1 e^{2z}+n_2}\right\}^{\frac{1}{2}(n_1+n_2)} e^{n_1 z}\,dz. \tag{15 1}
$$

We change to standard units $z = t\sigma_z + \bar{z}$, then

$$(15\,2)\qquad P(t) = \frac{1}{\sqrt{2\pi}}\left\{\frac{n_1+n_2}{n_1 e^{2(t\sigma+\bar{z})}+n_2}\right\}^{\frac{1}{2}(n_1+n_2)} e^{n_1(t\sigma+\bar{z})}\,dt, \qquad -\infty < t < \infty.$$

We rewrite this as

$$(15.3)\qquad P(t) = \frac{1}{\sqrt{2\pi}}\left\{\frac{n_1+n_2}{n_1 e^{2n_2(t\sigma+\bar{z})/(n_1+n_2)}+n_2 e^{-2n_1(t\sigma+\bar{z})/(n_1+n_2)}}\right\}^{\frac{1}{2}(n_1+n_2)} dt.$$

Expand $n_1 e^{2n_2(t\sigma+\bar{z})/(n_1+n_2)}$ and $n_2 e^{-2n_1(t\sigma+\bar{z})/(n_1+n_2)}$ and add term by term. Divide this result by n_1+n_2 from the numerator of $P(t)$ to obtain

$$(15.4)\qquad 1 + \frac{2n_1n_2(t\sigma+\bar{z})^2}{(n_1+n_2)^2} + O_1\left\{\frac{1}{(n_1+n_2)^{\frac{1}{2}}}\right\}.$$

Hence

$$(15.5)\qquad P(t) = \frac{1}{\sqrt{2\pi}}\left\{1 + \frac{2n_1n_2(t\sigma+\bar{z})^2}{(n_1+n_2)^2}\right\}^{-\frac{1}{2}(n_1+n_2)} dt.$$

We evaluate (15.5) for n_1 and n_2 large by using logarithms.

$$-\frac{n_1+n_2}{2}\log\left\{1+\frac{2n_1n_2(t\sigma+\bar{z})^2}{(n_1+n_2)^2}\right\}$$
$$= -\frac{n_1+n_2}{2}\left[\left\{\frac{2n_1n_2(t\sigma+\bar{z})^2}{(n_1+n_2)^2}\right\} - \frac{1}{2}\left\{\frac{2n_1n_2(t\sigma+\bar{z})^2}{(n_1+n_2)^2}\right\}^2 + \sum_{r=3}^{\infty}\frac{(-1)^{r+1}}{r}\left\{\frac{2n_1n_2(t\sigma+\bar{z})^2}{(n_1+n_2)^2}\right\}^r\right].$$

This gives

$$-\frac{\sigma^{-2}}{2}(t^2\sigma^2+2t\sigma\bar{z}+\bar{z}^2) + \frac{n_1^2n_2^2}{(n_1+n_2)^3}(t\sigma+\bar{z})^4 + \sum_{r=3}^{\infty}(-1)^r\frac{\{2n_1n_2(t\sigma+\bar{z})^2\}^r}{2r(n_1+n_2)^{2r-1}}.$$

We rèduce this then to

$$-\frac{t^2}{2} - \sigma^{-1}\bar{z}t - \frac{(\bar{z}\sigma^{-1})^2}{2} + \frac{1}{2}\left\{\frac{2n_1^2n_2^2}{(n_1+n_2)^2}\right\}\frac{(t\sigma+\bar{z})^4}{n_1+n_2}$$

+ terms involved in the above summation. Let $U = \sigma^{-1}\bar{z} < \sigma$. Since $\lim_{n_1,n_2\to\infty}\sigma = 0$, $\lim_{n_1,n_2\to\infty} U = 0$. Similarly $\lim_{n_1,n_2\to\infty}\frac{\bar{z}^2\sigma^{-2}}{2} = \lim_{n_1,n_2\to\infty}\frac{U^2}{2} = 0$. Consider $\frac{n_1^2n_2^2}{(n_1+n_2)^3}(t\sigma+\bar{z})^4 = \frac{\sigma^{-4}(t\sigma+\bar{z})^4}{4(n_1+n_2)} = \frac{(t+U)^4}{4(n_1+n_2)}$. Hence $\lim_{n_1,n_2\to\infty}\frac{(t+U)^4}{4(n_1+n_2)} = 0$. In like fashion

$$\sum_{r=3}^{\infty}\frac{(-1)^r}{2r}\left\{\frac{2n_1n_2}{n_1+n_2}\right\}^r\frac{(t\sigma+\bar{z})^{2r}}{(n_1+n_2)^{r-1}} = \sum_{r=3}^{\infty}\frac{(-1)^r\sigma^{-2r}(t\sigma+\bar{z})^{2r}}{2r(n_1+n_2)^{r-1}}.$$

Now clearly from our previous discussion for $r = 2$, we see

$$\lim_{n_1,n_2\to\infty} \sum_{r=3}^{\infty} \frac{(-1)^r}{2r} \frac{\sigma^{-2r}(t\sigma + \bar{z})^{2r}}{(n_1+n_2)^{r-1}} = 0.$$

This completes the proof.

We now consider the function, $f(z)$, determined by the approximate seminvariants of z. We start with

$$\lambda_{1:z} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) \quad \text{and} \quad \lambda_{r:z} = \frac{(r-2)!}{2}\left\{\frac{n_2+r-1}{n_2^r} + (-1)^r \frac{n_1+r-1}{n_1^r}\right\}, \quad r>1,$$

from (12 2) using only the first term. We may easily prove then that as n_1 and n_2 approach infinity in any manner whatever the function $f(z)$ represents a normal frequency distribution with

$$\bar{z} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) \quad \text{and} \quad \mu_{2:z} = \frac{1}{2}\left(\frac{n_2+1}{n_2^2} + \frac{n_1+1}{n_1^2}\right).$$

This further shows the identity of $f(z)$ and $y(z)$ in the limit as n_1 and $n_2 \to \infty$.

Since the moment generating function of $f(z)$ is

$$\left(1 - \frac{\theta}{n_2}\right)^{\frac{1}{2}(n_2-1-\theta)} \left(1 + \frac{\theta}{n_1}\right)^{\frac{1}{2}(n_1-1+\theta)}$$

we have

$$(15\,6) \qquad f(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\theta z}\left(1 - \frac{i\theta}{n_2}\right)^{\frac{1}{2}(n_2-1-i\theta)} \left(1 + \frac{i\theta}{n_1}\right)^{\frac{1}{2}(n_1-1+i\theta)} d\theta$$

I have not been able to evaluate (15 6) . We instead shall find an upper bound to the difference $|f(z) - y(z)|$ as n_1 and n_2 become large. We form $f(z) - y(z)$. Then by use of Stirling's formula for $n!$ with the remainder term and by the Fourier Integral Theorem,

$$(15.7) \quad |f(z) - y(z)| \leqq (e^{\beta_3/6n_1 + \beta_4/6n_2} - 1)y(z) \text{ where } 0<\beta_3<1, 0<\beta_4<1,$$

and

$$(15\,8) \qquad \lim_{n_1,n_2\to\infty} |f(z) - y(z)| = 0, \text{ and for this case } f(z) = y(z).$$

Of course (15.7) furnishes the upper bound of the absolute value between the frequency distribution of z and the function determined by the approximate seminvariants of z for any values of n_1 and n_2.

Up to this point we have assumed that there exists a function determined by the seminvariants

$$\lambda_{1:z} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) \quad \text{and} \quad \lambda_{r:z} = \frac{(r-2)!}{2}\left\{\frac{n_2+r+1}{n_2^r} + (-1)^r \frac{n_1+r-1}{n_1^r}\right\}, \quad r>1.$$

This may readily be proved by using the following theorem [18] (p. 536): The

determined character of the moments problem for an infinite interval is insured if $\sum_{n=1}^{\infty} c_{2n}^{-1/2n}$ diverges $\left(c_n = \int_{-\infty}^{\infty} x^n \, dF(x)\right)$.

16. The Pearson types of approximating curve. In discussing the types of the Pearson system which may be expected to approximate the z distribution we shall use the results of H. C. Carver [1], and the further exposition of C. C. Craig [3]. To find the Pearson type we compute $\delta = \dfrac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3}$. We shall find it convenient to use the approximations $\alpha_3 = \dfrac{\sqrt{2}\,(n_1 - n_2)}{\sqrt{n_1 n_2 (n_1 + n_2)}}$ and $\alpha_4 = 3 + 4\,\dfrac{(n_1^2 - n_1 n_2 + n_2^2)}{n_1 n_2 (n_1 + n_2)}$ to obtain

$$\delta = \frac{(n_1 + n_2)^2}{3n_1^2 n_2 + 3n_1 n_2^2 + 2n_1^2 - 2n_1 n_2 + 2n_2^2}, \tag{16.1}$$

and consequently $0 < \delta \leqq \frac{1}{2}$. The only possibilities are Types IV, VII, VI, or V since the greatest value of α_3^2 by (14.1) is 2.3565. Now if $n_1 = n_2$, we have Type VII, since $\alpha_3 = 0$, $\delta > 0$. In all other cases we shall have Types IV, V, or VI according as $\alpha_3^2 < 4\delta(\delta + 2)$, $\alpha_3^2 = 4\delta(\delta + 2)$, $\alpha_3^2 > 4\delta(\delta + 2)$. We neglect δ^2. Hence $\alpha_3^2 < 8\delta$ implies

$$\begin{aligned} &n_2^4(n_1 - 2) + n_2^3(15n_1^2 + 6n_1) + n_2^2(15n_1^3 - 8n_1^2) \\ &\qquad + n_2(n_1^4 + 6n_1^3) - 2n_1^4 > 0. \end{aligned} \tag{16.2}$$

A simple investigation reveals then the following results:

Type IV for $n_1, n_2 \geqq 2$, $n_1 \neq n_2$.

Type IV for $n_1 = 1$, $1 \leqq n_2 \leqq 21$; or $n_2 = 1$, $1 \leqq n_1 \leqq 21$.

(16.3) Type VI for $n_1 = 1$, $n_2 > 22$.

for $n_2 = 1$, $n_1 > 22$.

Type VII for $n_1 = n_2$.

Clearly the z distribution has features comparable to Type IV since both have infinite range. However, Type IV is irksome to fit in practice.

17. The Type III approximating curve, the logarithmic curve, and the Gram-Charlier Type A. The criterion for Type III is $\delta = 0$, $\alpha_3 \neq 0$. We see that as n_1 and n_2 increase the value of δ will decrease. Even for small values of n_1 and n_2 Type III will furnish a fair approximation to the z distribution. For example $n_1 = 10$, $n_2 = 5$, $\delta = .094$. The advantage of the Type III approxi-

mation rests on the fact that Salvosa's tables may be used. From the chart in [16] since $\alpha_3^2 \leqq 2\,3565$, we are assured that the approximating Type III curve is bell shaped For $n_1 = 1, 2$, $n_2 =$ any value, this approximation is not all that could be desired, although even in such cases it does have value. We note that Type III has limited range at one extreme $\left(-\frac{2}{\alpha_3}, \infty\right)$ while the range of the z distribution is $(-\infty, \infty)$. Salvosa's tables extend as far as $\alpha_3 = 1.1$, and since max $\alpha_3 = 1\,5351$, we see in some cases, and these only for $n_1 = 1$, n_2 large, we shall be obliged to make use of Pearson's *Tables of the Incomplete Gamma Function* [14] The logarithmic frequency curve

$$f(u) = \frac{1}{\sqrt{2\pi}\, c(u-a)} \exp\left[-\frac{1}{2c^2}\left(\log\frac{u-a}{b}\right)^2\right]$$

will be useful in approximating the z distribution While it has been discussed by many authors we shall follow Pae-Tsi Yuan [23], where a full bibliography may be found. In our discussion we use the $\beta_1 = \alpha_3^2$, $\beta_2 = \alpha_4$ chart of the Pearson system as given by S J Pretorius [16] (p. 147), since the logarithmic frequency locus connecting α_3^2 and α_4 is already drawn in The justification of this curve for fitting is due to the fact that in the β_1, β_2 chart of the Pearson system as given by S. J. Pretorius [16] (p 147), the logarithmic frequency locus lies in the Type VI region between the Type III locus and the Type V locus, and consequently closer to the Type IV region than Type III itself does. Hence since Type III fits fairly well under certain conditions and Type IV fits well we can expect the same for the logarithmic curve Furthermore when α_3 is small the logarithmic curve is similar to Type III [23] (p 42), and as α_3 becomes larger, $\alpha_3 = 1$, the difference between the two types is pronounced. However, it is just when α_3 becomes large in the region $n_1 = 1$, $n_2 \geqq 22$ that we find the logarithmic curves give a fine fit, since in such cases the point (α_3^2, β_2) lies practically on the logarithmic locus [16] To fit the curve [23] (pp 37, 48, 49), we find the values of the three parameters a, b, c To find c we solve the equation $w^3 + 3w^2 - (4 + \alpha_{3z}^2) = 0$ for w using the table [23] (p. 48) given by Pae-Tsi Yuan Knowing w we can easily solve for

$$c = (\log w)^{\frac{1}{2}}, \qquad b = \left(\frac{w+2}{\alpha_{3z}}\right) w^{-\frac{1}{2}} \sigma_z,$$

$$a = \bar{z} - \frac{(w+2)\sigma_z}{\alpha_{3z}}, \qquad t = \frac{z-\bar{z}}{\sigma_z} = \frac{e^{xc-\frac{1}{2}c^2} - 1}{(e^{c^2}-1)^{\frac{1}{2}}} \tag{17.1}$$

where the value of x must be obtained from the table of areas under the normal curve, if the z distribution is approximated by use of areas

Since the Gram-Charlier Type A series generally approximates a Pearson Type IV fairly well when α_3^2 is not too large, it is to be expected that the Type A series will approximate the z distribution in those cases when $n_1 = n_2$, and also when α_3^2 is not too large.

18. Levels of significance and approximation methods. We shall apply the results of the previous paragraphs to the determination of the value of z for any level of significance α, i.e. the value of z such that $\int_{-\infty}^{z} y(z)\,dz = 1 - \alpha$. We have such levels as the median (the 50% point of significance), the 20%, 5%, 1%, and .1% points as given in [9]. Where these tables apply there is no need for other methods. It would be desirable to extend the results for any level of significance whatever. The methods which we shall use are (1) the logarithmic frequency curve, (2) the Gram-Charlier Type A, and (3) the Type III approximation. For finding the levels of significance by the Incomplete Beta function, the reader is referred to [13], (p. lviii, topic (viii)). The logarithmic curve is very simple to use in conjunction with the table of areas under the normal curve. From Pae-Tsi Yuan we have

$$t = \frac{e^{xc-\frac{1}{2}c^2} - 1}{(e^{c^2} - 1)^{\frac{1}{2}}}, \qquad \text{where } (e^{c^2} - 1)^{\frac{1}{2}} \tag{18.1}$$

takes the same sign as α_3. The value of x is obtained from the table of the normal curve, 1.64 for the 5% level, 2.33 for the 1% level; the value of c is obtained from w (17.1), and consequently the value of t (18.1). Then we have if z_α = value of z for any level of significance, $t = \frac{z_\alpha - \bar{z}}{\sigma_z}$ to solve for z_α, where $\bar{z}$, and σ_z are the values of the mean and standard deviation of z as given by the proper formulas in (5), (6), (7). We illustrate with examples:

(18.2) 5% point of z, $n_1 = \infty$, $n_2 = 1$. $\alpha_3 = 1.5351$, $w = 1.2264$, $x = 1.64$, $t = 1.88$, $\bar{z} = .6352$, $\sigma_z = 1.11$, and as a result $z_{5\%} = 2.72$. Fisher [9] gives 2.7693.

We can also find $z_{5\%}$ easily for $n_1 = 1$, $n_2 = \infty$. Here $\alpha_3 = -1.5351$, $w = 1.2264$, $x = -1.64$, $t = 1.197$, $\bar{z} = -.6352$, $\sigma_z = 1.11$, $z_{5\%} = .694$ compared with Fisher [9] $z_{5\%} = .6729$.

(18.3) 1% point for $n_1 = 4$, $n_2 = 8$, $\bar{z} = -.0701$, $\sigma_z = .4819$, $\alpha_{3:z} = -.3619$, $w = 1.0144$, $t = 2.17$ and $z_{1\%} = .976$, while the accurate result is .9734.

From experience the values of z for any level of significance obtained by the logarithmic frequency curve will possess an error less than 2% of the true value of z for the level of significance if n_1 and n_2 are greater than twenty. It would seem that for other values of n_1 and n_2 the error could not be greater than 10%, and usually would be much less.

19. The Gram-Charlier Type A. We take the series in the form

$$F(t) = \varphi(t) + A_3\varphi^{(3)}(t) + A_4\varphi^{IV}(t), \qquad \varphi(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}}$$

$$t = \frac{z - \bar{z}}{\sigma_z}, \qquad A_3 = \frac{-\lambda_{3:t}}{3!}, \qquad A_4 = \frac{\lambda_{4:t}}{4!}.$$

Some examples follow.

(19.1) We use the material of (18.3) and employ three terms of $F(t)$. $\bar{z} = -.0701$, $\sigma_z = .4819$, $\lambda_{3\cdot z} = -.0405$, $\lambda_{4\cdot z} = 0336$, $A_3 = .06032$, $A_4 = .02596$.

Fitting $F(t)$ by ordinates we have $t = 2.17$, and consequently $z = .976$.

(19.2) We take $n_1 = n_2 = 5$, $\bar{z} = 0$, $\sigma_z = .4952$, $\lambda_{3\cdot z} = 0$, $\lambda_{4\cdot z} = 02798$, $A_3 = 0$, $A_4 = .01939$.

5% point: By ordinates $t = 1.57$, $z_{5\%} = .777$, while Fisher gives .8097.
1% point: By ordinates $t = 2.325$, $z_{1\%} = 1.15$, while Fisher gives 1.1974.

(19.3) We take $n_1 = 3$, $n_2 = 20$, $\bar{z} = -15909$, $\sigma_z = 5099$, $\lambda_{3\cdot z} = -.10222$, $\lambda_{4\cdot z} = 08822$, $A_3 = .12854$, $A_4 = .05438$ By ordinates $t = 1.523$, $z_{5\%} = .618$, Fisher gives .5654. $t = 1.989$, $z_{1\%} = .855$, Fisher gives .7985 The Gram-Charlier Type A is recommended only for $n_1 = n_2$ and n_1, $n_2 \geqq 20$

20. Type III approximation, the median, and 5% point. Since for Type III the median, m_z, is approximately two-thirds of the distance from the mode to the median if α_3 is moderate [12], [6], then we have further assuming n_1, $n_2 \geqq 20$

$$m_z = \frac{1}{3}\left(\frac{1}{n_1} - \frac{1}{n_2}\right) + \frac{1}{9}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right). \tag{20.1}$$

From experience this result will furnish an accuracy with an error less than 2% of the true value in the range above indicated.

$$t_{5\%} = 1.6437 + .2760\alpha_3 - .04506\alpha_3^2. \tag{20.2}$$

This was found by use of Salvosa's tables and for $\alpha_3 > 1\ 1$ by [14].

$$z_{5\%} = \sigma_z[1.644 + 2760\alpha_{3\cdot z} - .0451\alpha_{3\cdot z}^2] + \bar{z}. \tag{20.3}$$

We illustrate the use of (20.3) with some examples.

(20 4) $n_1 = n_2 = 1$, $\sigma_z = 1.5706$, $\alpha_{3,z} = 0$, $\bar{z} = 0$, $z_{5\%} = 2\ 582$,

while the accurate value is $z_{5\%} = 2.5421$.

(20 5) $n_1 = \infty$, $n_2 = 1$, $\alpha_3 = 1.5351$, $\bar{z} = .6352$, $\sigma_z = 1\ 11$, $z_{5\%} = 2\ 81$. The accurate value is 2.7693.

(20.6) $n_1 = n_2 = 5$, $\sigma_z = .4952$, $\alpha_{3\cdot z} = 0$, $\bar{z} = 0$, $z_{5\%} = .8141$, while the accurate value is $z_{5\%} = .8097$.

(20.7) $n_1 = 4$, $n_2 = 8$, $\bar{z} = -.0701$, $\sigma_z = .4819$, $\alpha_3 = -3619$, $z_{5\%} = .6712$, while the accurate value is .6725.

(20.8) $n_1 = 1$, $n_2 = 10$, $\bar{z} = -.5835$, $\sigma_z = 1.1353$, $\alpha_3 = -1.4333$, $z_{5\%} = .7283$, while the accurate value is .8012.

In a future paper exactly the same methods will be used for any per cent point of z whatever in order to compare with the results of W. G. Cochran [2] If

n_1 and n_2 are large we may use the approximate formulas for σ_z, $\alpha_{3:z}$, and $\bar{z}$ to obtain to the order of σ_z^3,

$$(20.9) \quad z_{5\%} = 1.644\sigma_z + .7760\left(\frac{1}{n_2} - \frac{1}{n_1}\right), \quad \text{where} \quad \sigma_z = \sqrt{\frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right)}.$$

We expand Fisher's result [9]
$z_{5\%} = \dfrac{1.6449}{\sqrt{n-1}} + .7843\left(\dfrac{1}{n_2} - \dfrac{1}{n_1}\right)$ by the binomial theorem, where $h = \dfrac{1}{\sigma_z^2}$, to obtain a comparable result

$$(20.10) \qquad z_{5\%} = 1.645\sigma_z + .7843\left(\frac{1}{n_2} - \frac{1}{n_1}\right).$$

The numerical examples given in this chapter illustrate unfavorable cases as well as favorable ones.

21. The distribution of F. Historically Snedecor [19] was the first to use F for e^{2z}. We find

$$(21.1) \qquad P(F) = \frac{n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \frac{F^{\frac{1}{2}n_1 - 1}}{(n_1 F + n_2)^{\frac{1}{2}(n_1 + n_2)}} dF, \qquad 0 \leqq F \leqq \infty.$$

The distribution of F is J shaped if $n_1 \leqq 2$, and bell shaped for $n_1 > 2$, and for $n_1 > 2$ one mode exists, $F_0 = \dfrac{n_2(n_1 - 2)}{n_1(n_2 + 2)}$. The two points of inflection, which exist for $n_1 \geqq 4$, are equidistant from the mode. The moments are

$$\mu_m' = \left(\frac{n_2}{n_1}\right)^m \frac{\Gamma\left(\frac{n_1 + 2m}{2}\right)\Gamma\left(\frac{n_2 - 2m}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}, \qquad n_2 > 2m$$

$$\bar{F} = \frac{n_2}{n_2 - 2}, \qquad n_2 > 2, \qquad \mu_2 = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)} \sim 2\left(\frac{1}{n_1} + \frac{1}{n_2}\right),$$

$$\alpha_{3:F} \sim \frac{2\sqrt{2}(2n_1 + n_2)}{\sqrt{n_1 n_2(n_1 + n_2)}}.$$

The exact results for μ_3, μ_4, α_3, and α_4 are omitted because of length. We have the theorem that as n_1, $n_2 \to \infty$ in any manner whatever the distribution of F approaches normality with mean $\bar{F} = 1$, $\sigma_F = \sqrt{2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$. The proof is omitted. The only type of approximating curve of any value is Type III. Of course the distribution of F is Type VI. No tables exist for Type VI. Furthermore the F distribution approaches the Type III function so slowly as to make most approximations of little value unless $\alpha_{3,F} \leqq 1.1$. Other possible

parameters are $\theta = \frac{n_1(n_2+1)}{n_2(n_1+1)}F$, and $H = \frac{n_1 F}{n_2 + n_1 F}$, [13] Since $|\alpha_{3\cdot H}| = 2|\alpha_{3\cdot z}|$ approximately we see that the distribution of H is more skewed than that of z. We mention briefly also $S_1^2 - S_2^2$ where $S_1^2 = \frac{n_1}{N_1}s_1^2$, $S_2^2 = \frac{n_2}{N_2}s_2^2$. Clearly z, F, θ, and H give equivalent levels of significance This is not true for z and $S_1^2 - S_2^2$.

Finally, since $F = \frac{s_1^2}{s_2^2}$, it may be interpreted as a quotient [5]. When the moments of F do not exist, it is due to the distribution function of s_2^2

22. Conclusion. We have found the seminvariants for the z distribution, and approximations for them Type III, and the logarithmic normal frequency functions are shown to be excellent approximations to the z distribution The approach to normality for the z distribution is proved A formula is given for finding the 5% level of significance for z. The F distribution is studied along the same lines As far as the construction of tables for levels of significance is concerned, the z distribution is much easier to use My sincerest thanks are due Professor C C Craig for his helpful guidance and many suggestions.

BIBLIOGRAPHY

[1] H C Carver, *Handbook of Mathematical Statistics*, H L Rietz, ed , Boston Houghton-Mifflin Co , 1924 Chapter on frequency curves

[2] W. G Cochran, "Note on an approximate formula for the significance levels of z," *Annals of Math Stat* , Vol 11 (1940), pp 93–95

[3] C. C. Craig, "A new exposition and chart for the Pearson system of frequency curves," *Annals of Math. Stat.*, Vol. 7 (1936), pp 16–28

[4] C. C. Craig, "An application of Thiele's semi-invariants to the sampling problem," *Metron*, Vol 7 (1928–29), pp. 3–74

[5] C C. Craig, "The frequency function of y/x," *Annals of Math* , Second Series, Vol. 30 (1929), pp 471–486

[6] A T Doodson, "Relation of a mode, median, and mean in a frequency curve," *Biometrika*, Vol. 11 (1917), p 425

[7] R. A Fisher, "On a distribution yielding the error functions of several well known statistics," *Proc International Math. Cong* , 1924, Toronto, Vol 2, pp. 805–813

[8] R A Fisher, and E A Cornish, "Moments and cumulants in the specification of distributions," *Revue de l'Institut International de Statistics*, 5th year, pp. 307–20, 1937, La Hague

[9] R A Fisher, and Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, London· Oliver and Boyd, 1938.

[10] K. Knopp, *Theory and Application of Infinite Series*, English translation, Edinburgh: Blackie and Son, 1928

[11] N Nielsen, *Handbuch der Theorie der Gamma Functionen*, Leipzig Teubner, 1906

[12] C. A. Olshen, "Transformation of the Pearson Type III distribution," *Annals of Math Stat.*, Vol. 9 (1938), pp. 176–200.

[13] K. Pearson (Editor), *Tables of the Incomplete Beta Function*, London. Biometrika Office, University College, London, 1934

[14] K. Pearson (Editor), *Tables of the Incomplete Gamma Function*, London. His Majesty's Stationery Office, 1922.

[15] K. Pearson, S. A Stouffer, and F. N. David, "Further applications in statistics of the $T_n(x)$ Bessel function," *Biometrika*, Vol. 24 (1932), pp. 293 350.

[16] S. J. Pretorius, "Skew bivariate frequency curves examined in the light of numerical illustrations," *Biometrika*, Vol. 22, (1930-31).

[17] L. R. Salvosa, "Tables of Pearson's Type III function," *Annals of Math. Stat.*, Vol. 1 (1930), pp 191-8 et seq.

[18] J. Shohat, and M Frechet, "A proof of the generalized second limit-theorem in the theory of probability," *Trans. Am. Math Soc.*, Vol. 33 (1931), pp. 531 43.

[19] G. W Snedecor, *Calculation and Interpretation of the Analysis of Variance and Covariance*, Ames, Iowa. Collegiate Press.

[20] T. J Stieltjes, "Tables des valeurs des sommes $s_k = \sum_{n=1}^{\infty} n^{-k}$," *Acta Math.*, Vol 10, pp 299 302.

[21] Whittaker and Robinson, *The Calculus of Observations*, Edinburgh: Blackie and Son, second edition, p. 135

[22] Whittaker and Watson, *Modern Analysis*, 4th edition, London: Cambridge University Press, 1935.

[23] Pae-Tsi Yuan, "On the logarithmic frequency distribution and the semi-logarithmic correlation surface," *Annals of Math. Stat.*, Vol. 4, (1933).

[24] R. T. Zoch, "Some interesting features of frequency curves," *Annals of Math Stat.*, Vol. 4, (1935), pp. 1-10.

THE DOOLITTLE TECHNIQUE

By Paul S. Dwyer

University of Michigan

1. Introduction. Most authors who have presented the Doolittle method, from Doolittle [1] down to the present, have not given any formal proof that the solution is valid in the general case. They usually are content with a form describing the various steps of a Doolittle solution.

The author has recently shown [2] that the Doolittle method can be abbreviated to a technique which is also an abbreviation, essentially, of the method of single division and its abbreviation which Aitken called the "Method of Pivotal Condensation" [3]. It appears at once that the validity of the Doolittle method follows from the validity of the method of single division—a validity which is readily established.

However one may desire a "proof" which is based directly on the Doolittle technique without referring to other methods of solution. It is the chief purpose of this paper to present such a proof. It is accomplished by the introduction of a notation which precisely describes the conventional Doolittle process and by proving that this process results in a system of equations whose prediagonal terms are zero. It is a secondary purpose of the paper to emphasize the advantages of the Abbreviated Doolittle method and to explain and illustrate minor variations in the conventional Doolittle technique.

2. The Abbreviated Doolittle solution. We first direct our attention to the essential parts of a Doolittle solution and these are the last two rows of each matrix of the standard Doolittle presentation. The additional rows in the standard presentation are rows of products which are used solely for the purpose of finding the two bottom rows of each matrix and they need not be recorded, if a computing machine is available, since the essential information is present in the two bottom rows. Doolittle [1] did not have calculating machines (he used multiplication tables) but he put the important information in Table A and carefully segregated the supplementary information in Table B. With reference to this he wrote [1]

"It is to be observed that the numbers in Table B have but a single use while those in Table A are used over and over, and where the number of equations is large, it is of great advantage that they should be thus tabulated by themselves in a form compact and easy of reference."

For purposes of proof, as well as for purposes of calculation if a computing machine is available, it is only necessary to utilize the forward part of the Abbreviated Doolittle solution which is the equivalent of the Doolittle Table A.

A four variable illustration of the Abbreviated Doolittle technique is presented in Table I. The successive equations are indicated by number, as is customary, and the operation which defines the equation is specified. The actual operation is indicated more explicitly by the notation of column 3 and this is discussed in the next section

The presentation of Table I introduces one variation from the standard Doolittle method The division is made by the diagonal coefficient of each row rather than by its negative. One may still use the old technique, if he prefers, but it is felt that one can subtract products as easily as he can add products with modern machines equipped with automatic negative multiplication. In addition the entries of the equivalent rows then have the same signs and, too, it is not necessary to take the time to change the signs of the second rows. This variation uses the same division method as the method of single division [2] and as the method of pivotal condensation [3] so that the abbreviated form of these methods is, essentially, the same as the abbreviated form of the Doohttle method.

The application of this technique leads at each step to a coefficient for each variable. However if the process is to lead from our four equations in four unknowns, to three in three, to two in two, to one in one, it follows that all the entries to the left of the diagonal, which we may call prediagonal entries, must be zero. That this is true in the general case is the objective of the proofs of later sections.

3. A notation for and description of the Doolittle technique. A main contribution of the present article is the use of a notation which describes the Doolittle technique. As long as the Doolittle process is described loosely by means of "operations" it is difficult to be precise in defining quantities which appear in the calculation, but when a notation is used which is definite enough to permit expansion in terms of the original coefficients, some sort of proof may be available. The present notation bears some resemblance to that suggested by Gauss [4], though Gauss used letters to indicate the primary subscripts and numbers to indicate the number of secondary subscripts and his notation was directly applicable to the sums of least squares theory rather than to symmetric equations in general.

We wish to find the solution of the equations

$$\sum_{i=1}^{n} a_{ij} x_i = a_{n+1,j}, \qquad j = 1, 2, \cdots, n \tag{1}$$

where the matrix of the coefficients is symmetric. We do this by obtaining auxiliary equations which feature a decreasing number of variables. No serious restriction is made if we assume that the variables x_1, x_2, x_3, etc., are eliminated successively. The Doolittle technique may then be described as follows:
We take the first equation of (1) and divide by its leading coefficient, a_{11}, to get

TABLE I

Abbreviated Doolittle technique, forward solution

Eq	Operation	Notation	x_1	x_2	x_3	x_4	
I		a_{i1}	a_{11}	a_{21}	a_{31}	a_{41}	a_{51}
II		a_{i2}	a_{12}	a_{22}	a_{32}	a_{42}	a_{52}
III		a_{i3}	a_{13}	a_{23}	a_{33}	a_{43}	a_{53}
IV		a_{i4}	a_{14}	a_{24}	a_{34}	a_{44}	a_{54}
V	I repeated	a_{i1}	a_{11}	a_{21}	a_{31}	a_{41}	a_{51}
VI	V divided by a_{11}	$b_{i1} = \dfrac{a_{i1}}{a_{11}}$	1	b_{21}	b_{31}	b_{41}	b_{51}
VII	II—b_{21} V	$a_{i2\cdot1} = a_{i2} - a_{i1}b_{21}$	$a_{12\cdot1}$	$a_{22\cdot1}$	$a_{32\cdot1}$	$a_{42\cdot1}$	$a_{52\cdot1}$
VIII	VII divided by $a_{22\cdot1}$	$b_{i2\cdot1} = \dfrac{a_{i2\cdot1}}{a_{22\cdot1}}$	$b_{12\cdot1}$	1	$b_{32\cdot1}$	$b_{42\cdot1}$	$b_{52\cdot1}$
IX	III—b_{31} (V)—$b_{32\cdot1}$ VII	$a_{i3\cdot12} = a_{i3} - a_{i1}b_{31} - a_{i2\cdot1}b_{32\cdot1}$	$a_{13\cdot12}$	$a_{23\cdot12}$	$a_{33\cdot12}$	$a_{43\cdot12}$	$a_{53\cdot12}$
X	IX divided by $a_{33\cdot12}$	$b_{i3\cdot12} = \dfrac{a_{i3\cdot12}}{a_{33\cdot12}}$	$b_{13\cdot12}$	$b_{23\cdot12}$	1	$b_{43\cdot12}$	$b_{53\cdot12}$
XI	IV—b_{41} V—$b_{42\cdot1}$ VII—$b_{43\cdot12}$ IX	$a_{i4\cdot123} = a_{i4} - a_{i1}b_{41} - a_{i2\cdot1}b_{42\cdot1} - a_{i3\cdot12}b_{43\cdot12}$	$a_{14\cdot123}$	$a_{24\cdot123}$	$a_{34\cdot123}$	$a_{44\cdot123}$	$a_{54\cdot123}$
XII	XI divided by $a_{44\cdot123}$	$b_{i4\cdot123} = \dfrac{a_{i4\cdot123}}{a_{44\cdot123}}$	$b_{14\cdot123}$	$b_{24\cdot123}$	$b_{34\cdot123}$	1	$b_{54\cdot123}$

$$\sum_{i=1}^{n} b_{i1}x_i = b_{n+1,1}, \qquad \text{where} \quad b_{i1} = \frac{a_{i1}}{a_{11}}, \tag{2}$$

and we then form

$$\sum_{i=1}^{n} a_{i2\cdot 1}x_i = a_{n+1,2\cdot 1} \qquad \text{with} \quad a_{i2\cdot 1} = a_{i2} - a_{i1}b_{21}. \tag{3}$$

We then divide by $a_{22\cdot 1}$ and get

$$\sum_{i=1}^{n} b_{i2\cdot 1}x_i = b_{n+1,2\cdot 1} \qquad \text{with} \quad b_{i2\cdot 1} = \frac{a_{i2\cdot 1}}{a_{22\cdot 1}}. \tag{4}$$

We next form

$$\sum_{i=1}^{n} a_{i3\cdot 12}x_i = a_{n+1,3\cdot 12} \qquad \text{with} \quad a_{i3\cdot 12} = a_{i3} - a_{i1}b_{31} - a_{i2\cdot 1}b_{32\cdot 1}, \tag{5}$$

and

$$\sum_{i=1}^{n} b_{i3\cdot 12}x_i = b_{n+1,3\cdot 12} \qquad \text{with} \quad b_{i3\cdot 12} = \frac{a_{i3\cdot 12}}{a_{33\cdot 12}}. \tag{6}$$

This process is continued so that, in general, we have

$$\sum_{i=1}^{n} a_{ij\cdot 12\cdots j-1}x_i = a_{n+1,j\cdot 12\cdots j-1}, \qquad j = 1, 2, \cdots, n \tag{7}$$

and

$$\sum_{i=1}^{n} b_{ij\cdot 12\cdots j-1}x_i = b_{n+1,j\cdot 12\cdots j-1}, \qquad j = 1, 2, \cdots, n \tag{8}$$

with

$$\begin{aligned} a_{ij\cdot 12\cdots j-1} = a_{ij} &- a_{i1}b_{j1} - a_{i2\cdot 1}b_{j2\cdot 1} - a_{i3\cdot 12}b_{j3\cdot 12} - \cdots \\ &- a_{i,j-2\cdot 12\cdots j-3}b_{j,j-2\cdot 12\cdots j-3} - a_{i,j-1\cdot 12\cdots j-2}b_{j,j-1\cdot 12\cdots j-2} \end{aligned} \tag{9}$$

and

$$b_{ij\cdot 12\cdots j-1} = \frac{a_{ij\cdot 12\cdots j-1}}{a_{jj\cdot 12\cdots j-1}}. \tag{10}$$

It is to be noted that the n equations (1) are transformed by this process to the n auxiliary equations of (7) or (8). The solutions of (1) are also solutions of these auxiliary equations since the auxiliary equations are linear combinations of (1). It is our purpose to show that the prediagonal coefficients of these auxiliary equations are always 0 so that these auxiliary equations feature a decreasing number of variables.

We may use the term primary subscripts to indicate the first two subscripts and the term secondary subscripts to indicate the later subscripts which specify the order of elimination of the variables. The "order" of the coefficient is then equal to the number of secondary subscripts.

The formula (9) gives the matrix of the final Doolittle set of equations. At each stage of the reduction one can write down a formula for all the elements in the matrix at that stage Thus one can write the coefficients of order h, $a_{ij\cdot 12\ \cdots h}$, in terms of coefficients of order less than h,

$$a_{ij\cdot 12\ \cdots h} = a_{ij} - a_{i1}b_{j1} - a_{i2\ 1}b_{j2\ 1} - \cdots - a_{i,h-1\cdot 12\ \cdots h-2}b_{j,h-1\cdot 12\ \cdots h-2} - a_{ih\cdot 12\ \cdots h-1}b_{jh\cdot 12\ \cdots h-1}. \tag{11}$$

It follows at once that

$$\begin{aligned} a_{ij\cdot 12\ \cdots h} &= a_{ij\cdot 12\ \cdots h-1} - a_{ih\cdot 12\ \cdots h-1}b_{jh\ 12\ \cdots h-1} \\ &= a_{ij\cdot 12\ \cdots h-1} - \frac{a_{ih\ 12\ \cdots h-1}a_{jh\cdot 12\cdots h-1}}{a_{hh\ 12\ \cdots h-1}}. \end{aligned} \tag{12}$$

4. Some theorems on the interchangeability of subscripts. Our main objective is to prove that the prediagonal terms are zero. In order to do this we first prove some theorems dealing with the primary and secondary subscripts.

THEOREM 1. *The value of* $a_{ij\cdot\ \cdots h}$ *is not changed if the primary subscripts are interchanged* This theorem which might be stated "The matrix of the coefficients of a given order is symmetric" follows from the symmetry of the matrix of coefficients of zero order We can show that the symmetry of the matrix having coefficient of order h follows at once from the symmetry of the matrix having coefficients of order $h - 1$ by comparing the value $a_{ij\cdot\ \cdots h}$ with that of $a_{ji\cdot\ \cdots h}$ obtained by dual substitution in (12). Since the matrix of zero order coefficients is symmetric by hypothesis, it follows that the matrices of the coefficients of order 1, 2, 3, 4, etc., are in turn symmetric.

THEOREM 2: *Any pair of consecutive secondary subscripts may be interchanged without changing the value of the coefficient* This theorem indicates that, within prescribed limits, the order of elimination does not have any effect on the result

Consider the coefficient $a_{ij\cdot\cdots kl\cdots}$ having r secondary subscripts before the k and s secondary subscripts after the l and consider the corresponding coefficient $a_{ij\cdot\cdots lk\cdots}$ which results from an interchange of k and l These coefficients can be expressed by continued use of (12) in terms of coefficients of order $r + 2$. The resulting expansion of $a_{ij\cdot\cdots kl\cdots}$ is equivalent to that of $a_{ij\cdot\ \cdots lk\ \cdots}$ with the interchange of the l and the k. It follows that the theorem is true if $a_{ij\cdot\cdots lk\cdots} = a_{ij\cdot\cdots kl\cdots}$. Now a double application of (12) to $a_{ij\cdot\cdots lk\cdots}$ leads to the expansion in terms of coefficients of order r (using the notation $a_{ij\cdot}$ to indicate the coefficient of the r-th order)

$$a_{ij\cdot\ \cdots kl\cdots} = a_{ij\cdot} - \frac{a_{ik\cdot}a_{jk\cdot}}{a_{kk\cdot}} - \frac{\left(a_{il\cdot} - \dfrac{a_{ik\cdot}a_{lk\cdot}}{a_{kk\cdot}}\right)\left(a_{jl\cdot} - \dfrac{a_{jk\cdot}a_{lk\cdot}}{a_{kk\cdot}}\right)}{a_{ll\cdot} - \dfrac{a_{lk\cdot}^2}{a_{kk\cdot}}}. \tag{13}$$

Then $a_{ij\cdot\cdots lk\cdots}$ is expanded similarly, the difference is formed and found to be zero.

It follows that the theorem is true

The application of Theorem 2 with the continued interchange of successive secondary subscripts in all possible ways leads at once to

THEOREM 3: *The secondary subscripts may be interchanged in all possible ways without changing the value of the coefficient.* This theorem might be stated "The value of the resulting coefficient is independent of the order of elimination." This is the sort of result one would expect to find and indeed, some may feel that it is intuitively evident, but this formal proof is presented for those who desire a more rigorous approach.

Theorem 3 enables us to prove Theorem 4 which may be stated: *The value of* $a_{ij\cdot 12\ldots n}$ *is always zero if at least one of the secondary subscripts is equal to one of the primary subscripts.*

Suppose i is this subscript. Then by Theorem 3, i may be placed in the final position. Now by (12) we have

$$a_{ij,\ldots,i} = a_{ij,\ldots} - \frac{a_{ij,\ldots}\, a_{ii,\ldots}}{a_{ii,\ldots}} = 0.$$

A similar statement holds if j appears among the secondary subscripts.

5. The vanishing of the prediagonal entries. As an application of Theorem 4 we can show that the prediagonal entries are identically zero and this is exactly what is needed to establish the validity of the forward Doolittle process. It is to be noted that the prediagonal entries are of form $a_{i,j,12\ldots,j-1}$ with $i < j$. Then i must equal one of the secondary subscripts and the term is zero.

It follows that no entries need be made to the left of the diagonal in the Abbreviated Doolittle solution and, indeed, no entries need be made in the original matrix below the main diagonal. A numerical problem is presented in the next section.

6. Illustration. The Abbreviated Doolittle technique is illustrated in Table II. This illustration is essentially an illustration of a previous article [2] and serves as the basis, in a later section, for expansion into the standard Doolittle solution. The check is shown in the right hand column and the back solution is indicated. The check entries for the first matrix are obtained by adding the entries in the row to the main diagonal and then adding the entries in the column. All other check entries are obtained by adding the entries in the row.

The solution is easily made once it is understood and results from continued application of formula (9). For example

$$a_{54\cdot 123} = a_{54} - a_{51}b_{41} - a_{52\cdot 1}b_{42\cdot 1} - a_{53\cdot 12}b_{43\cdot 12}$$

and this is

$$a_{54\cdot 123} = .8000 - (.2000)(.6000) - (.3200)(.1905) - (.4619)(-.1612) = .6935$$

(see the underscored entries of Table II). Terms of this sort are easily computed if a calculating machine, and especially so if one equipped with automatic

positive and negative multiplication, is available The back solution too is easily accomplished with a machine It is only necessary to substitute in turn in each of the "b" equations. Thus the value of x_1 is $\frac{a_{54\cdot 123}}{a_{44\cdot 123}} = b_{54\ 123}$, the value of x_2 is $b_{53\cdot 12} - b_{43\ 12}b_{54\cdot 123} = b_{53\cdot 124}$, that of x_3 is $b_{52\ 1} - b_{42\ 1}b_{54\cdot 123} - b_{32\ 1}b_{53\cdot 124} = b_{52\cdot 134}$, etc. The back solution of the check is treated similarly.

7. A variation in technique. Before proceeding with the presentation of a standard Doolittle solution it seems wise to indicate another possible variation in the technique in addition to the division by the diagonal coefficient rather than its negative. It is possible to obtain the Doolittle solution by using the fixed entry from the first of the equivalent rows in place of using the fixed "b" entry and the variable "a". This results from the fact that

$$a_{ik}\ldots\ b_{jk}\ldots = a_{jk}\ldots\ b_{ik}\ldots \quad \left(= \frac{a_{ik}\ldots\ a_{jk}\ldots}{a_{kk}\ldots}\right). \tag{14}$$

Thus in Table II the value $a_{54\cdot 123}$ can be obtained with the use of

$$a_{54\ 123} = a_{54} - a_{41}b_{51} - a_{42\ 1}b_{52\ 1} - a_{43\cdot 12}b_{53\ 12}$$

as readily as with the use of

$$a_{54\cdot 123} = a_{54} - a_{51}b_{41} - a_{52\ 1}b_{42\ 1} - a_{53\ 12}b_{43\cdot 12}$$

See the boxed entries of Table II.

There seems to be no real choice between these techniques. The fixed "b" is traditional in the standard Doolittle solution while the abbreviation of the method of single division leads to a fixed "a". The point to be emphasized here is that either the fixed "a" or the fixed "b" can be used. Also (14) is used in the next section in supplying details for the check portion of a standard Doolittle method.

8. The standard Doolittle method. If no computing machine is available or if a more detailed solution is desired, it is preferable to record the individual products of (9) and thus arrive at the standard Doolittle method. (The division by the diagonal coefficient rather than its negative is not a fundamental difference.) The standard Doolittle method, from this point of view, is an expanded form of the Abbreviated Doolittle method with more details added. Its validity then follows from the validity of the Abbreviated Doolittle method While it is not true that all prediagonal terms vanish in the standard Doolittle method, and this fact complicates the check by row sums, yet the prediagonal a_{ij}. . (and $b_{ij}\ldots$) are all zero.

The standard Doolittle method is presented in Table III. Some remarks should be made about the non-recorded terms, the two check solutions, and the back solution.

The blanks (—) indicate non zero entries which are usually not presented in a

Doolittle solution. They should be considered however if the first check method is to be used.

The first check method, which is the logical extension of the check method of the Abbreviated Doolittle solution, has been outlined by Ezekiel [5]. The row sum is the sum of all the entries in the row whether recorded or not. In order to check, it is necessary to add these unrecorded entries, and they are available

TABLE II

Abbreviated Doolittle Solution; illustration

x_1	x_2	x_3	x_4		Check
1.0000	.4000	.5000	.6000	.2000	2.7000
—	1.0000	.3000	.4000	.4000	2.5000
—	—	1.0000	.2000	.6000	2.6000
—	—	—	1.0000	.8000	3.0000
1.0000	.40000	.5000	[.6000]	.2000	2.7000
1.0000	.40000	.5000	.6000	[.2000]	2.7000
	.8400	.1000	[.1600]	.3200	1.4200
	1.0000	.1190	.1905	[.3810]	1.6905
		.7381	[−.1190]	.4619	1.0810
		1.0000	−.1612	[.6258]	1.4646
			.5903	.6935	1.2837
			1.0000	1.1748	2.1747
		1.0000		.8152	1.8152
	1.0000			.0602	1.0602
1.0000				−.9366	.0635

in the columns above if we make use of formula (12). Thus, if we wish to check the value $\sum_{i=1}^{5} a_{i1}b_{41} = 1.6200$, we have

$$a_{11}b_{41} + a_{21}b_{41} + a_{31}b_{41} + a_{41}b_{41} + a_{51}b_{41} =$$

$$a_{41} + a_{41}b_{21} + a_{41}b_{31} + a_{41}b_{41} + a_{51}b_{41} =$$

$$.6000 + .2400 + .3000 + .3600 + .1200 = 1.6200.$$

Another check method, which is recommended by Peters and Van Voorhis [6] sums the entries in the row only over those columns which are to be recorded.

This is presented as check method 2 of Table III. As is to be expected, the check values of the a's and b's of the last two rows of each matrix are in agreement.

It might be noted that one may use the first check method without checking the intermediate steps (the sums for each row) if he checks the sums for the last two rows of each matrix.

TABLE III

Doolittle solution, with checks

Notation	x_1	x_2	x_3	x_4		Check Method 1	Check Method 2	
a_{i1}	1 0000	4000	5000	6000	2000	2 7000	2.7000	
a_{i2}	—	1 0000	3000	.4000	4000	2 5000	2.1000	
a_{i3}	—	—	1 0000	.2000	6000	2 6000	1 8000	
a_{i4}	—	—	—	1 0000	8000	3 0000	1 8000	
a_{i1}	1 0000	.4000	5000	.6000	2000	2 7000	2 7000	
b_{i1}	1 0000	4000	5000	.6000	2000	2 7000	2 7000	
a_{i2}	—	1 0000	.3000	4000	4000	2.5000	2 1000	
$a_{i1}b_{21}$	—	1600	2000	.2400	0800	1 0800	6800	
$a_{i2\cdot1}$		8400	1000	.1600	3200	1 4200	1.4200	
$b_{i2\cdot1}$		1 0000	1190	.1905	.3810	1 6905	1 6905	
a_{i3}	—	—	1 0000	2000	6000	2.6000	1 8000	
$a_{i1}b_{31}$	—	—	2500	3000	1000	1 3500	6500	
$a_{i2\cdot1}b_{32\ 1}$		—	.0119	0190	0381	1690	0690	
$a_{i3\cdot12}$			.7381	−.1190	.4619	1 0810	1 0810	
$b_{i3\cdot12}$			1.0000	−.1612	6258	1 4646	1.4646	
a_{i4}	—	—	—	1 0000	.8000	3.0000	1 8000	
$a_{i1}b_{41}$	—	—	—	3600	1200	1 6200	4800	
$a_{i2\ 1}b_{42\cdot1}$		—	—	0305	0610	2705	0914	
$a_{i3\ 12}b_{43\ 12}$			—	0192	− 0745	− 1743	− 0553	
$a_{i4\cdot123}$				5903	6935	1 2838	1 2839	
$b_{i4\ 123}$				1.0000	1.1748	2 1748		
$b_{i3\ 124}$			1 0000	− 1894	8152	1 81532	− 3506	
$b_{i2\cdot134}$		1 0000	0970	2238	0602	1.0602	4143	2160
$b_{i1\cdot234}$	1 0000	0241	4076	.7049	− 9366	.0634	1 3049	.9076 4241

The back solution is carried out as in Table II. If no computing machine is available or if the detailed steps are desired they may be indicated as in Table III. The entries in the box under the x_4 column are respectively $b_{54\cdot123}b_{43\ 12}$, $b_{54\cdot123}b_{42\cdot1}$, and $b_{54\cdot123}b_{41}$. Those in the preceding column are $b_{53\ 124}b_{32\ 1}$ and $b_{53\ 124}b_{31}$. The other entry is $b_{52\ 134}b_{21}$. The values of the coefficients are obtained by subtracting these row entries from the constant term of the corresponding "b" equation. Thus, $b_{53\ 124}$ = (.6258) − (−.1894); $b_{52\cdot134}$ =

(.3810) − .0970 − .2238, etc. The back solution of check method 1 agrees with that of check method 2. A form for accomplishing the back solution of the check is indicated at the right. It is not necessary to complete the back solution of the check if it is not desired, and indeed, there are some who feel that the use of the row sum check is unnecessary with modern computing machines [7]. The basic check is substitution in the original equations.

9. Summary. The chief purpose of this paper is to show that the Doolittle technique actually leads to a set of equations featuring a decreasing number of unknowns. This is accomplished by the introduction of an appropriate notation to describe the process and the establishment of certain theorems which serve to validate the process. These theorems are of some interest aside from the application made here. It is a secondary purpose of this paper to emphasize the practicability and theoretical advantages (relative ease of calculating, theoretically more accurate, less chance for numerical error, less recording, less time consuming, more compact, and more easily checked) of the Abbreviated Doolittle method and to explain and illustrate possible variations in technique in the forward and check (by row sums) portions of the standard Doolittle solution. It should be noted that the notation suggested is very useful in providing an easy development of various theorems used in multiple and partial correlation studies, the presentation of which is not the purpose of the present paper.

REFERENCES

[1] M. H. Doolittle, "Method employed in the solution of normal equations and the adjustment of a triangulation," *U. S. Coast and Geodetic Survey Report* (1878), pp. 115–120.

[2] P. S. Dwyer, "The solution of simultaneous equations," *Psychometrika*, Vol. 6 (1941), pp. 101–129.

[3] A. C. Aitken, "Studies in practical mathematics I. The evaluation, with applications, of a certain triple product matrix," *Roy. Soc. Edin. Proc.*, Vol. 57 (1937), pp. 172–181.

[4] C. F. Gauss, "Supplementum theoriae combinationis observationum erroribus minimis obnoxiae," *Werke*, Vol. 4 (1873), pp. 60–71.

[5] Mordecai Ezekial, *Methods of Correlation Analysis*. John Wiley and Sons, Inc., New York (1930), pp. 362–364.

[6] C. C. Peters, and W. R. Van Voorhis, *Statistical Procedures and Their Mathematical Bases*, McGraw Hill (1940), pp. 228–229.

[7] A. K. Kurtz, "The use of the Doolittle method in obtaining related multiple correlation coefficients," *Psychometrika*, Vol. 1 (1936), pp. 45–51.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

A PROBLEM IN ESTIMATION

By Joseph F Daly

The Catholic University of America

Several recent psychological studies in the field of memory testing [1], [2], [3] have suggested the following problem Let each individual E in our population be characterized by the variates $y^1, \cdots, y^p; y^{p+1}, \cdots, y^{p+t}$ $(p > t)$ Suppose, however, that circumstances make it impossible for us to observe the last t variates. For example, we may think of $y^1, \cdots, y^p$ as an individual's scores on a battery of tests, and think of $y^{p+1}, \cdots, y^{p+t}$ as measures of certain psychological characteristics which, though affecting the individual's performance, are not subject to direct observation. To make up for this, assume that we have a theory which tells us that if $y^{p+1}, \cdots, y^{p+t}$ are held constant, then the observable y's are dependent upon them according to a specified regression equation

$$y^i = x^i_\mu y^\mu, \qquad (i = 1, \cdots, p; \mu = p+1, \cdots, p+t).$$

Somewhat more precisely, we assume the distribution laws

$$(1) \quad f(y^1, \cdots, y^{p+t}) = (2\pi)^{-\frac{1}{2}(p+t)} \mid A_{rs} \mid^{\frac{1}{2}} \exp\{-\tfrac{1}{2}A_{rs}(y^r - a^r)(y^s - a^s)\},$$

(where $r, s = 1, \cdots, p+t$, and repeated indices are to be summed according to the usual convention) and

$$(2) \quad f(y^1, \cdots, y^p \mid y^{p+1}, \cdots, y^{p+t}) = (2\pi\sigma^2)^{-\frac{1}{2}p} \exp\left\{-\frac{1}{2\sigma^2}\sum_i (y^i - x^i_\mu y^\mu)^2\right\}.$$

The x^i_μ are supposed to be known, but except for the conditions imposed by (1) and (2) nothing is known about the quantities A_{rs}, a^r, and σ^2. Having observed the test scores y^i_α $(\alpha = 1, \cdots, N)$ obtained by N individuals E_α drawn at random from the population, we wish to estimate the values $y^{p+1}_\alpha, \cdots, y^{p+t}_\alpha$ corresponding to each E_α, and the essential parameters in the distribution law (1), particularly the variances and covariances of $y^{p+1}, \cdots, y^{p+t}$.

We can easily find optimum estimates of the y^μ_α by applying the method of maximum likelihood to the function (2) after substituting for the y^i the scores y^i_α obtained by the individual in question. Thus if we write

$$v_{\mu\nu} = x^i_\mu x^i_\nu, \qquad \| v^{\mu\nu} \| = \| v_{\mu\nu} \|^{-1},$$

(assuming thereby that the rank of the matrix $\| x^i_\mu \|$ is l) we have

$$\hat{y}^\mu_\alpha = v^{\mu\nu} x^i_\nu y^i_\alpha . \tag{3}$$

These estimates are unbiased in the sense that the expected value of $\hat{y}^\mu$ calculated from the distribution law (2) is y^μ.

But when we come to estimate the variances and covariances involved in (1), the procedure is less straightforward. Under the present circumstances we cannot use the expression

$$\frac{1}{N-1} \sum_\alpha (y^\mu_\alpha - \bar{y}^\mu)(y^\nu_\alpha - \bar{y}^\nu), \tag{4}$$

for the sample covariance of y^μ and y^ν. We might, of course, try substituting the estimates $\hat{y}^\mu_\alpha$ from (3) for the unknown y^μ_α in (4). But this expedient will in general produce a biased estimate. Denoting the required covariance by $A^{\mu\nu}$ (the element in the appropriate position in the inverse of the matrix $\| A_{\rho\sigma} \|$), we find as a matter of fact that the expected value of (4) when the y^μ_α are replaced by their estimates $\hat{y}^\mu_\alpha$ is

$$A^{\mu\nu} + \sigma^2 v^{\mu\nu}. \tag{5}$$

This bias may or may not be important in any given case. But it can conceivably be quite serious if the $A^{\mu\nu}$ are relatively small, especially if such expressions are employed in the usual way to estimate the correlation coefficient rather than the covariance.

Perhaps the most logical way to attack the problem is through the joint distribution of $y^1, \cdots, y^p$ alone, obtainable by integrating the undesirable variates $y^{p+1}, \cdots, y^{p+l}$ out of (1). We therefore consider

$$f(y^1, \cdots, y^p) = (2\pi)^{-\frac{1}{2}p} \, | \bar{A}_{ij} |^{\frac{1}{2}} \exp \{ -\tfrac{1}{2} \bar{A}_{ij} (y^i - a^i)(y^j - a^j) \}, \tag{6}$$

where

$$\bar{A}_{ij} = A_{ij} - A_{i\mu} B^{\mu\nu} A_{\nu j}, \qquad \| B^{\mu\nu} \| = \| A_{\mu\nu} \|^{-1}.$$

Moreover, when account is taken of (2), we find that we must have

$$A_{ij} = \frac{\delta_{ij}}{\sigma^2} \qquad A_{i\mu} = -\frac{x^i_\mu}{\sigma^2} \qquad a^i = x^i_\mu a^\mu$$

(δ_{ij} being Kronecker's delta). If we now form the likelihood function $\prod_{\alpha=1}^{N} f(y^1_\alpha, \cdots, y^p_\alpha)$ from (6) for our sample, and set its derivatives with respect to the a^μ, σ^2, and the $B^{\mu\nu}$, equal to zero, we arrive, after some simplification, at the equations

$$a^\mu = v^{\mu\nu} x^i_\nu \bar{y}^i = \frac{1}{N} \sum \hat{y}^\mu_\alpha, \qquad \text{[cf. (3)]}$$

$$
\begin{aligned}
&\left\{A^{ij} - \frac{1}{N}\sum_{\alpha}(y^i_\alpha - x^i_\mu a^\mu)(y^j_\alpha - x^j_\nu a^\nu)\right\}\delta_{ij} = 0,\\
&\left\{A^{ij} - \frac{1}{N}\sum_{\alpha}(y^i_\alpha - x^i_\mu a^\mu)(y^j_\alpha - x^j_\nu a^\nu)\right\}x^i_\sigma x^j_\tau = 0,\\
&A^{ij} = \sigma^2\delta^{ij} + x^i_\mu A^{\mu\nu}x^j_\nu,
\end{aligned}
\tag{7}
$$

for determining the maximum likelihood estimates. The first of equations (7) is already solved for the a^μ, and the solution of the simultaneous equations for the remaining essential parameters yields the estimates

$$\hat{\sigma}^2 = \frac{1}{N(p-l)}\sum_{\alpha,i}(y^i_\alpha - x^i_\mu \hat{y}^\mu_\alpha)^2 \tag{8}$$

$$\hat{A}^{\mu\nu} = \frac{1}{N}\sum_{\alpha}(\hat{y}^\mu_\alpha - \hat{a}^\mu)(\hat{y}^\nu_\alpha - \hat{a}^\nu) - v^{\mu\nu}\hat{\sigma}^2. \tag{9}$$

A considerable amount of algebraic manipulation is required to put the solutions in the form given above, but since the results are about what one would expect in view of (5), we omit the details. As is often the case, some bias remains in the "optimum" estimates (9). However, this can be eliminated by writing $N - 1$ in place of N The estimate (8) of σ^2 is unbiased as it stands.

REFERENCES

[1] T. V. Moore, "The analysis of association by its equational constants," *Aspects of the New Scholastic Philosophy*, New York, Benziger Brothers, 1932, pp 181–225

[2] Regis Holland, "The development of logical and rote memory," *Studies in Psychology and Psychiatry*, the Catholic University of America, Vol 4, no 8 (1940).

[3] Helen deSales Forrest, "Correlations between the constants in the curve of learning," *Studies in Psychology and Psychiatry*, the Catholic University of America, Vol. 5, no. 1 (1941)

CONFIDENCE LIMITS FOR AN UNKNOWN DISTRIBUTION FUNCTION

By A. Kolmogoroff

Moscow, U.S.S.R.

Let $x_1, x_2, \cdots, x_n$ be mutually independent random variables following the same distribution law

$$P\{x_i \leq \xi\} = F(\xi). \tag{1}$$

A recent paper by A. Wald and J. Wolfowitz[1] deals with the problem of using

[1] A. Wald and J. Wolfowitz, "Confidence limits for distribution functions," *Annals of Math. Stat*, Vol 10 (1939), pp. 105–118

the observable values of the x's to estimate the function $F(\xi)$. In this connection it may be useful to recall the following results published by me in 1933.[2]

Put

$$F_n(\xi) = \frac{N(\xi)}{n} \tag{2}$$

where $N(\xi)$ denotes the number of those x's whose observed values do not exceed ξ.

THEOREM 1: *If the function $F(\xi)$ is continuous then the distribution law of the quantities*

$$D_n = \sup | F(\xi) - F_n(\xi) | \sqrt{n} \tag{3}$$

does not depend on $F(\xi)$.

Denote by $\Phi_n(\lambda)$ the value of the probability $P\{D_n \leq \lambda\}$ which is common to all continuous distribution functions $F(\xi)$.

THEOREM 2: *For n tending to infinity, the distribution function $\Phi_n(\lambda)$ tends to*

$$\Phi(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2} \tag{4}$$

uniformly with respect to λ.

A more elementary proof of Theorem 2 was given by N. Smirnoff in 1939.[3] Another paper by the same author[4] gives a table of the function $\Phi(\lambda)$.

Without the assumption that $F(\xi)$ is continuous, we easily obtain

THEOREM 3: *Whatever be the distribution function $F(\xi)$,*

$$P\{D_n \leq \lambda\} \geq \Phi_n(\lambda). \tag{5}$$

Theorems 1 and 3 giving the exact lower bound of the probability that $F_n(\xi)$ will satisfy the inequality

$$| F(\xi) - F_n(\xi) | \leq \frac{\lambda}{\sqrt{n}} \tag{6}$$

for all values of ξ, can be used to establish confidence limits for $F(\xi)$ corresponding to the confidence coefficient

$$\alpha = \Phi_n(\lambda). \tag{7}$$

These confidence limits will be free from any restriction concerning the nature of the function $F(\xi)$.

---

[2] A. Kolmogoroff, "Sulla determinatione empirica di una legge di distributione," *Giornale dell'Istituto Italiano degli Attuari*, Vol. 4 (1933), pp. 83-91.

[3] N Smirnoff, "Sur les écarts de la courbe de distribution empirique," *Recueil Math. de Moscou*, Vol. 6 (1939), pp. 3-26.

[4] N. Smirnoff, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bulletin de l'Université de Moscou, Série internationale (Mathématiques)*, Vol. 2, fasc. 2 (1939).

For sufficiently large values of n we can use the limiting distribution (4) and write

$$\alpha = \Phi(\lambda). \tag{8}$$

The following short table, based on that of Smirnoff,[4] gives the values of λ corresponding to a few chosen confidence coefficients α.

TABLE OF λ

α	λ
.95	1.35
.98	1.52
.99	1.63
.995	1.73
.998	1.86
.999	1.95

Smirnoff's paper[4] contains still another application of the function $\Phi(\lambda)$ Denote by $x_1', x_2', \cdots, x_{n_1}'$ and $x_1'', x_2'', \cdots, x_{n_2}''$ two sequences of mutually independent random variables following the same probability law $F(\xi)$. Let further $F_{n_1}(\xi)$ and $F_{n_2}(\xi)$ be two random step functions corresponding to these series, defined as in (2). Smirnoff proves then the following

THEOREM 4: *If the probability law $F(\xi)$ is continuous, then the probability*

$$P\left\{\sup \mid F_{n_1}(\xi) - F_{n_2}(\xi) \mid \leq \lambda \sqrt{\frac{n_1 + n_2}{n_1 n_2}}\right\} = \Phi_{n_1, n_2}(\lambda) \tag{9}$$

is independent of the function $F(\xi)$. If n_1 and n_2 are indefinitely increased subject to the restriction that the ratio n_1/n_2 remains between two fixed numbers a_1 and a_2

$$0 < a_1 \leq \frac{n_1}{n_2} \leq a_2 < +\infty \tag{10}$$

then

$$\Phi_{n_1, n_2}(\lambda) \to \Phi(\lambda). \tag{11}$$

In the general case, where the probability law $F(\xi)$ is absolutely arbitrary we have

$$P\left\{\sup \mid F_{n_1}(\xi) - F_{n_2}(\xi) \mid \leq \lambda \sqrt{\frac{n_1 + n_2}{n_1 n_2}}\right\} \leq \Phi_{n_1, n_2}(\lambda). \tag{12}$$

Owing to the above results the quantity

$$D_{n_1, n_2} = \sup \mid F_{n_1}(\xi) - F_{n_2}(\xi) \mid \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \tag{13}$$

could be used as a criterion to test the hypothesis that the probability laws of the two series of observable variables are actually the same.

CORRECTIONS TO A PAPER ON THE UNIQUENESS PROBLEM OF MOMENTS

By M. G. Kendall

London, England

I wish to make certain corrections in my paper on "Conditions for Uniqueness in the Problem of Moments" (*Annals of Math. Stat.*, Vol. II (1940), p. 402). I thought I had succeeded in improving on results given earlier by Stieltjes, Lévy and Carleman, but this is not so.

Theorem 1 of the paper stated that a set of moments determines a distribution uniquely if $\sum_{r=0}^{\infty} \frac{\nu_r t^r}{r!}$ converges for some real non-zero t, ν_r being the absolute moment of order r. This is true, and a similar result has been proved by Lévy, but my proof contained a small lacuna. It was shown that the characteristic function $\phi(t)$ has a Taylor expansion which, under the conditions of the theorem, is convergent; but it has also to be shown that it is equal to the sum of that expansion. This may be seen as follows:

We have

$$e^{itx} - \sum_{r=0}^{n-1} \frac{(itx)^r}{r!} \leq \theta \frac{|tx|^n}{n!}, \qquad |\theta| \leq 1,$$

and hence, on taking mean values,

$$\left| \phi(t) - \sum_{r=0}^{n-1} \frac{(it)^r \mu_r}{r!} \right| \leq \frac{\nu_n t^n}{n!}.$$

Since by hypothesis $\frac{\nu_n t^n}{n!} \to 0$, $\phi(t)$ must be equal to the sum of its (convergent) Taylor expansion.

The principal error was a statement that $\nu_n^{1/n}/n$ must either tend to a limit or diverge. For this reason, the second theorem should run: a distribution determines a distribution uniquely if $\lim \nu_n^{1/n}/n$ is finite (not $\lim \nu_n^{1/n}$, n as originally stated). Theorem 3 should also be restated with the upper limit substituted for the limit therein.

Theorem 4 stated that a set of moments uniquely determines a distribution if $\Sigma \frac{1}{\nu_n^{1/n}}$ diverges. A rigorous proof is as follows:

The characteristic function obeys the relation

$$|\phi^{(n)}(t)| \leq \nu_n, \qquad n > 1$$

provided, of course, that ν_n exists. A theorem of Denjoy[1] states that if a function $f(x)$, defined in the segment (a, b), possesses derivatives of all orders therein,

[1] Arnaud Denjoy, "Sur les fonctions quasi-analytiques de variable réelle," *Comptes Rendus* Vol 173 (1921), p. 1399.

if M_n is the maximum of $|f^{(n)}(x)|$ in the segment and if $\Sigma \frac{1}{M_n^{1/n}}$ is divergent, then $f(x)$ is completely determined by its value and that of its derivatives at a single point. $\phi(t)$ obeys the conditions of the theorem and by taking the point to be $t = 0$, theorem 4 follows.

I hope that this note will correct any misunderstandings that may have arisen on the main paper, and I regret that a number of circumstances, not the least of which is war, have made it impossible to forward the correction at an earlier date.

ANNOUNCEMENT CONCERNING COMPUTATION OF MATHEMATICAL TABLES

In the December, 1939, issue of the *Annals of Mathematical Statistics*, p. 399, there appeared an Announcement of the *Mathematical Tables Project*. This project is operated by the Work Projects Administration of New York City, as O. P No. 265–2–97–11 under the technical supervision of Dr. A. N. Lowan. It is sponsored by the National Bureau of Standards, Dr. Lyman J. Briggs, Director.

In order to keep the readers of the *Annals* up-to-date on the progress of the work of the Project, information will be released from time to time.

The following list shows the status of work, as of October, 1941. The reader is referred to the December, 1939 issue of the *Annals* with respect to which n will denote the n^{th} item of Tables Published, Pn will denote the n^{th} item of Tables in Progress and Cn will denote the n^{th} item of Tables under Consideration.

Tables published. 1, 2, 3, *P*1, *P*2, *P*3, *P*4, *P*6(b), *P*6(c), *P*6(d), *P*6(e), *P*7, *C*7 and also

1. Table of Five-Point Lagrangian Interpolants for arguments ranging between 0 and 2 at intervals of 0.001.

2. Tables of Grid Coordinates (American Polyconic Projection) at 5 minute intervals of latitude and longitude for latitude from 70°N to 28°N and for latitude from 49°N to 72°N.

3. Table for Map Projections of Northwestern Extension of U. S.

Tables in process of reproduction. *P*5, *P*6(a), *P*8 and *C*1 for [0 (.001) 7 (.01) 50 (.1) 300 (1) 2,000 (10) 10,000; 12*D*] also

1. Tables of Section Moduli and Moments of Inertia for Structural Members used in Naval Architecture. (For the Bureau of Marine Inspection and Navigation.)

2. Tables of $Si(x)$ and $Ci(x)$ for x ranging from 10 to 100 at intervals of 0.001.

3. The zeros of the Legendre Polynomials up to the 16th order to 15 decimal places and the Weight Coefficients for Gauss' Mechanical Quadrature Formula

Tables for which manuscripts are completed. $P9$, $P11$, $C6$, (the function x^y, instead of $A(x, y)$, has been tabulated to 15 places), and also

1 Table of $\int_0^x J_0(t)\,dt$ from 0 to 10 at intervals of 0.01 to 10 places.

Tables for which computations are completed. $P10$ (also $\tanh x$, $\coth x$), $C2$, $C3$, (change to $n = -21, -20 \cdots 0$) and also

1. Various hydraulic tables based on Kutter's and Manning's formulae. (Tabulation suggested by the War Department)

2. Table of reciprocals of the integers from 100,000 to 200,000.

3. Table of the Associated Legendre Functions $P_n^m(x)$ and $Q_n^m(x)$ for n ranging between 1 and 10, and m between 0 and 4; for arguments x and ix where x ranges between 0 and 10 at intervals of 0.1. Also corresponding values for half-integral values of n and values of the functions for arguments in degrees. (Tabulation suggested by National Defense Research Committee.)

4. Tables of $R \sin \theta$ and $R \cos \theta$ $R = 1000\ (10)\ 10{,}000$, $\theta = 5(5)800$ (in mils).

Tables for which computations are in progress. $C3$ (for $n = 1, 2, \cdots 20$) and also

1. Table of the Bessel Functions $Y_0(z)$ and $Y_1(z)$ for the same complex arguments as in $J_0(z)$ and $J_1(z)$, mentioned in $P9$.

2. Tables of Length of Meridional Arc at one-minute intervals.

3. Tables of the Confluent Hypergeometric Function for selected values of the parameters.

4. Tables of three-point, four-point, six-point and seven-point Lagrangian Interpolants

5. Table of Tchebysheff Polynomials.

Tables under consideration. $C4$ and also

1. Table of the first 10 powers of the reciprocals of the integers from 1 to 1,000.

2. Extensive tables of Elliptic Functions for both real and imaginary arguments.

3. A 12-place table of Inverse Circular and Hyperbolic Functions other than Arc tan x.

4. Table of the Integral $\int_0^x Y_0(t)\,dt$.

5. Tables of the non-periodic solutions of the Mathieu Differential Equation.

6. Table of the Error Functions for complex arguments (suggested by Federal Communication Commission).

7. Tables of the Unit-Sigma Functions and their integrals.

8. Tables of Circular Functions for Complex Arguments.

9. Tables of the Zeros of the Hermite and Laguerre Polynomials and of the corresponding Weight Factors in Gauss' Mechanical Quadrature Formula.

10. Table of Lamé Polynomials.

11. Table of Military Grid Coordinates for certain "Control Stations." (For the War Department.)

12. Tables of the Chi-Square Distribution and "Student's" *t*-distribution.

13. Tabulation of Fisher's *A*-, *B*-, and *C*- Distributions of the Multiple Correlation Coefficients.

The Project would welcome suggestions for the computation of new tables of interest in pure and applied mathematics, as well as information regarding computational work in progress elsewhere.

Communications should be addressed to Major Irving V. Huie, Administrator, Work Projects Administration, 70 Columbus Avenue, New York City.

Requests for copies of published tables should be addressed to Dr. Lyman J. Briggs, Director of the National Bureau of Standards, Washington, D. C.

REPORT OF THE CHICAGO MEETING OF THE INSTITUTE

The Fourth Summer Meeting of the Institute of Mathematical Statistics was held at The University of Chicago, Tuesday to Thursday, September 2 to 4, 1941, in conjunction with the meetings of the American Mathematical Society, the Mathematical Association of America, and the Econometric Society. The following sixty-eight members of the Institute attended the meeting.

R. L. Anderson, T. W. Anderson, K. J. Arnold, H. M. Bacon, Walter Barsky, W. D. Baten, A. A. Bennett, Paul Boschan, I. W. Burr, J. H. Bushey, W. F. Colerberg, W. G. Cochran, A. T. Craig, C. C. Craig, J. H. Curtiss, J. F. Daly, W. E. Deming, J. L. Doob, P. L. Dressel, P. S. Dwyer, Churchill Eisenhart, M. L. Ekeblaw, H. P. Evans, J. Fischer, W. C. Flaherty, R. M. Foster, C. H. Graves, Louis Guttman, W. I. Hare, F. C. Hinds, A. S. Householder, E. V. Huntington, William Hurwitz, M. H. Ingraham, Donham Jackson, Leo Katz, J. F. Kenney, L. A. Knowler, L. F. Knudsen, Tjalling Koopmans, C. F. Kossack, O. E. Lancaster, D. H. Leavens, B. A. Lengyel, W. G. Madow, J. N. Mehr, A. M. Mood, J. E. Morton, Leah Naugle, Harold Nisselson, J. I. Northam, E. G. Olds, Oystein Ore, C. K. Payne, G. A. D. Preinreich, Francis Regan, Selby Robinson, C. I. Rose, M. M. Sandomire, Max Sasuly, Henry Scheffe, H. M. Schwartz, Harry Silber, J. H. Smith, M. E. Wescott, S. S. Wilks, E. W. Wilson, Gale Young.

The opening session, on Tuesday morning, was devoted to contributed papers on *Probability and Statistics* and was held jointly with the American Mathematical Society and the Econometric Society. The Chairman was Professor A. T. Craig, University of Iowa, and the following papers were presented:

1. *A geometric derivation of Fisher's z-transformation.*
   J. B. Coleman, University of South Carolina.
2. *Large sample distribution of the likelihood ratio.*
   Abraham Wald, Columbia University.
3. *On the integral equation of renewal theory.*
   (Read by title.)
   Willy Feller, Brown University.
4. *Cumulative frequency functions.*
   Irving Burr, Purdue University.
5. *On spherical probability distributions.*
   K. J. Arnold, Massachusetts Institute of Technology.
6. *Some observations on analysis of variance theory.*
   (Read by title.)
   Hilda Geiringer, Bryn Mawr College.
7. *On the asymptotic distribution of medians of samples from a multivariate population.*
   A. M. Mood, University of Texas.
8. *A problem of estimation.*
   J. F. Daly, Catholic University.

Abstracts of these papers follow this report.

On Tuesday afternoon a session was held jointly with the Econometric Society on *Time Series Analysis*. Under the chairmanship of Professor C. C. Craig of the University of Michigan, the following papers were presented:

1. *Is sampling theory applicable to economic time series?*
   Tjalling Koopmans, Penn Mutual Life Insurance Co., Philadelphia
2. *Serial correlation*
   R. L. Anderson, North Carolina State College

The morning session on Wednesday was held jointly with the Econometric Society on *Curve Fitting*. The chair was held by Dr. J. Marschak of the New School for Social Research and the following papers were presented:

1. *Weights to compensate for transformation in curve fitting*
   T. C. Yntema, University of Chicago and Cowles Commission
2. *Curve fitting by cumulative addition*
   John H. Smith, University of Chicago and Cowles Commission

On Wednesday afternoon, Professor S. S. Wilks of Princeton University acted as chairman of a session on *Multivariate Analysis*. The following papers were read:

1. *On testing sets of means and discriminant analysis*
   Abraham Wald, Columbia University
2. *On tests of hypotheses concerning variances and covariances*
   William G. Madow, Bureau of the Census

The Josiah Willard Gibbs Lecture of the American Mathematical Society was delivered on Wednesday evening by Professor Sewall Wright of the University of Chicago. His topic was *Statistical Genetics and Evolution*.

On Thursday morning a joint session on *Demand and Supply Analysis* was held with the Econometric Society. At this session Dr. C. F. Roos of the Institute of Applied Econometrics presided, and the following papers were presented:

1. *Demand analysis for certain commodities based on income and budget data*
   J. Marschak, New School for Social Research, and George Garvy, National Bureau of Economic Research.
2. *Derivation of elasticities of demand and supply. A direct method*
   Oscar Lange, University of Chicago and Cowles Commission
3. *On the workings of a general equilibrium system*
   J. L. Mosak, University of Chicago and Cowles Commission

An informal reception was held on Monday evening in the Judson Court Lounge. On Tuesday and Wednesday afternoons the ladies of the Mathematics Department of the University of Chicago served tea in the Eckhart Hall Common Room. After the joint session on Tuesday afternoon, the Cowles Commission for Research in Economics gave a tea in the Common Room of the Science Building. On Thursday evening a joint dinner of the four mathematical organizations was held in Hutchinson Commons, preceded by an informal reception at the Reynolds Club.

EDWIN G. OLDS,
Secretary

ABSTRACTS OF PAPERS

(Presented on September 2, 1941, at the Chicago Meeting of the Institute)

A Geometric Derivation of Fisher's z-transformation. J. B. COLEMAN, University of South Carolina.

In fitting points in a plane by a line so that the sum of the squares of the perpendicular deviations shall be a minimum, a second line is found for which the sum of the squares of the deviations is a maximum. Let Σd^2 be the sum of the squares of the deviations of the points from the minimum line, and ΣD^2 be the sum of the squares from the maximum line. Then $\Sigma D^2/\Sigma d^2 = (1+r)/(1-r)$. $\frac{1}{2}\log(1+r)/(1-r)$ is Fisher's z-transformation for testing the coefficient of correlation.

Large Sample Distribution of the Likelihood Ratio. ABRAHAM WALD, Columbia University.

The large sample distribution of the likelihood ratio has been derived by S. S. Wilks (*Annals of Math Stat.*, Vol. 9 (1938)) in case of a linear composite hypothesis and under the assumption that the hypothesis to be tested is true. Here a general composite hypothesis is considered and the distribution in question is derived also in case that the hypothesis to be tested is not true. Let $f(x_1, \cdots, x_p, \theta_1, \cdots, \theta_k)$ be the joint probability density function of the variates $x_1, \cdots, x_p$ involving k unknown parameters $\theta_1, \cdots, \theta_k$. Denote by H_ω the hypothesis that the true parameter point $\theta = (\theta_1, \cdots, \theta_k)$ satisfies the equations $\xi_1(\theta) = \cdots = \xi_r(\theta) = 0$, $(r \leq k)$. Denote by λ_n the likelihood ratio statistic for testing H_ω on the basis of n independent observations on $x_1, \cdots, x_p$. For any parameter point θ let $\xi_{ij}(\theta) = \dfrac{\partial \xi_i(\theta)}{\partial \theta_j}$ and let $c_{ij}(\theta)$ be the expected value of $\dfrac{\partial \log f(x_1, \cdots, x_p, \theta)}{\partial \theta_i} \cdot \dfrac{\partial \log f(x_1, \cdots, x_p, \theta)}{\partial \theta_j}$ calculated under the assumption that θ is the true parameter point. For any θ denote by $A(\theta)$ the matrix $\| \xi_{ij}(\theta) \|$ $(i = 1, \cdots, r; j = 1, \cdots, k)$ and let $\| \sigma_{ij}(\theta) \| = \| c_{ij}(\theta) \|^{-1}$, $(i, j = 1, \cdots, k)$. Let furthermore $\| \sigma^*_{uv}(\theta) \|$, $(u, v = 1, \cdots, r)$ be the matrix equal to the product $A(\theta) \cdot \| \sigma_{ij}(\theta) \| \cdot \bar{A}(\theta)$, where $\bar{A}(\theta)$ is the tranapose of $A(\theta)$. Finally let $\| \bar{c}^*_{uv}(\theta) \| = \| \sigma^*_{uv}(\theta) \|^{-1}$, $(u, v = 1, \cdots, r)$. For each n and θ denote by $y_{1n}(\theta), \cdots, y_{rn}(\theta)$ a set of r variates which have a joint normal distribution with mean values $\sqrt{n}\xi_1(\theta), \cdots, \sqrt{n}\xi_r(\theta)$ and covariance matrix $\| \sigma^*_{uv}(\theta) \|$, $(u, v = 1, \cdots, r)$. Denote the quadratic form $\sum_{u=1}^{r} \sum_{v=1}^{r} y_{un}(\theta) y_{vn}(\theta) \bar{c}^*_{uv}(\theta)$ by $Q_n(\theta)$. It has been shown that under certain assumptions on $f(x_1, \cdots, x_p, \theta)$, $\xi_1(\theta), \cdots, \xi_r(\theta)$ we have $\lim_{n=\infty} |P(-2 \log \lambda_n < t \mid \theta) - P[Q_n(\theta) < t \mid \theta]| = 0$ uniformly in t and θ, where for any z $P(z < t \mid \theta)$ denotes the probability that $z < t$ holds under the assumption that θ is the true parameter point. The distribution of $Q_n(\theta)$ is known and has been treated in the literature. If H_ω is true, then $\xi_1(\theta) = \cdots = \xi_r(\theta) = 0$, and $Q_n(\theta)$ has the χ^2 distribution with r degrees of freedom.

On the Integral Equation of Renewal Theory. W. FELLER, Brown University.

As is well-known, the equation $U(t) = G(t) + \int_0^t U(t-x)\, dF(x)$ has frequently been

discussed, under different forms, in connection with the population theory, the theory of industrial replacement, etc. In the present paper it is shown that, using Tauberian theorems for Laplace integrals, it becomes possible to analyze in detail the asymptotic behavior of $l(t)$ as $t \to \infty$ and also to solve some other problems which have been discussed in the literature. Strict conditions for the validity of different methods to treat the equation are given together with some modifications found to be necessary. The paper will appear in the *Annals of Mathematical Statistics*.

Cumulative Frequency Functions. I. W. Burr, Purdue University.

Frequency and probability functions play a fundamental role in statistical theory and practice. They are, however, often inconvenient and difficult to use, since it is necessary to integrate or sum to find the probability for a given range. Theoretically the cumulative or integral frequency function would seem to be better adapted to determining such probabilities, since the latter can be found simply by a subtraction. The aim of this paper is to make a contribution toward the direct use of cumulative frequency functions. Some general properties and theory of cumulative functions are presented with particular emphasis upon certain moment functions adapted to such direct use. Both continuous and discrete cases are included. A list of possible cumulative functions is given and a particular one, $F(x) = 1 - (1 + x^c)^{-k}$, discussed fully. This function has properties which make it practicable and adaptable to a wide variety of distribution types. It well illustrates the possibilities of the cumulative approach.

On Spherical Probability Distributions. Kenneth J. Arnold, Massachusetts Institute of Technology.

Two methods of correspondence for circular distributions to the normal error function have led to non-constant absolutely continuous functions [See F. Zernike's article in *Handbuch der Physik* Vol. 3, pp. 477–478]. The corresponding distributions for the sphere are found. The case of diametrical symmetry for both circle and sphere is discussed. Tables of the probability integrals involved are given and an application in geology is included.

Some Observations on Analysis of Variance Theory. Hilda Geiringer, Bryn Mawr College.

The test functions used in analysis of variance present themselves in different classes of important problems. Their distribution has been determined and tabulated by *R. A. Fisher*[1] under the hypothesis that the chance variables are all *independent* of each other and subject to the *same normal* law. Consequently we can in this way test only the hypothesis that the theoretical populations have all these properties.

If it is not possible to determine the exact distribution of test functions under sufficiently general assumptions regarding the populations we may: (a) find an asymptotic solution of the problem, i.e. determine the distribution of the test functions *for large samples*.[2] Or (b) determine at least the mathematical expectations and the variances of the test functions for *appropriately general* populations and for *small samples*.

It is well known that the expectations of the two quadratic forms which are basic in the analysis of variance are *equal*, even if the n populations are not normal but equal to each other (Bernoulli series). But, in addition, we can prove the mathematical theorem that, under the same conditions the *expectation of their quotient equals one*. The next step consists in studying the case that the n distributions are not equal to each other and to investigate certain *inequalities* characteristic for the Lexis Series and Poisson Series. These different criteria are completed by the *computation of the variances* of the test functions.

[1] "Metron," Vol. 5 (1926), p. 90–104.

[2] See e.g. W. G. Madow, *Annals of Math. Stat.*, Vol. 11 (1940), p. 193.

In addition to the above mentioned test functions known as "variance within" and "variance among" classes other *symmetrical* test functions have been considered in the classical analysis of variance. Here again we may assume quite *general populations*. It results that the Lexis as well as the Poisson Series may now be characterized by *equalities* (instead of inequalities).

Finally it seems to be worthwhile to omit the assumption of independent chance variables and to study different kinds of *mutual* dependence. These investigations lead to new instructive *inequalities among the expectations*. These last considerations seem to be connected with Fisher's "intraclass correlation" and to supplement this idea.

On the Asymptotic Distribution of Medians of Samples from a Multivariate Population. A. M. Mood, University of Texas.

Let two variates x_1 and x_2 have a density function $f(x_1, x_2)$ which, besides being positive or zero and having its integral over the whole space equal to one, shall satisfy these conditions:

$$\int_{-\infty}^{\infty} f\left(x_1, \frac{1}{n}\right) dx_1 = \int_{-\infty}^{\infty} f(x_1, 0)\, dx_1 + O\left(\frac{1}{n}\right)$$

$$\int_{-\infty}^{\infty} f\left(\frac{1}{n}, x_2\right) dx_2 = \int_{-\infty}^{\infty} f(0, x_2)\, dx + O\left(\frac{1}{n}\right)$$

The coordinate system is assumed to have been chosen so that the population median is at the origin. Let $(\tilde{x}_1, \tilde{x}_2)$ be the median of a sample of $2n + 1$ elements drawn from a population with this density function. It is shown that for large samples $(\tilde{x}_1, \tilde{x}_2)$ is normally distributed to within terms of order $1/\sqrt{n}$ with zero means and variances and covariances given by certain integrals of $f(x_1, x_2)$.

A similar result is true for k as well as two variates.

A Problem in Estimation, Joseph F. Daly, The Catholic University of America.

Consider a normal population in which each individual is characterized by the variates $y_1, \cdots, y_p, y_{p+1}, y_{p+2}$. Suppose that the latter two are not directly observable, but that for given values of y_{p+1}, y_{p+2} the first set of y's is independently distributed about the "regression line" $y_k = y_{p+1} + k y_{p+2}$ $(k = 1, \cdots, p)$ with a common variance σ^2. For each individual, one can thus determine values $\hat{y}_{p+1}, \hat{y}_{p+2}$ from the observed $y_1, \cdots, y_p$, using the method of least squares. Assuming a similar relation between the expected values of $y_1, \cdots, y_{p+2}$ in the original population, these estimates $\hat{y}_{p+1}, \hat{y}_{p+2}$ are, of course, unbiased. However, if we calculate these $\hat{y}$'s for each individual of a sample of N, and substitute them in the Pearson product-moment correlation formula, the estimate of the correlation between y_{p+1} and y_{p+2} thus obtained is somewhat biased. The bias depends on the number of observable y's, and on the size of the variances and covariances of y_{p+1}, y_{p+2} relative to σ^2.

Is Sampling Theory Applicable to Economic Time Series? T. J. Koopmans, Penn Mutual Life Insurance Company.

The classical regression theory assumes that the values of the independent variables remain the same in repeated samples. Certain situations in economic analysis, like price formation according to the "cobweb" theorem, require a sampling theory of serial regression in which certain observations may represent a dependent variable at one time and an independent variable at a later time. This leads to the problem of the joint distribution of certain quadratic forms in normal variables.

The simplest problem of this type is that of the distribution of the ratio $r = q/p$ of a quadratic form q in T observations from a normal distribution with mean 0 to the sum p

of the squares of these observations. The distribution of r is independent of that of η and is

$$h(r) = \frac{\frac{1}{2}T - 1}{2\pi i} \int_\gamma \frac{(z - r)^{\frac{1}{2}T - 2}}{\left\{\prod_{l=1}^{T} (z - k_l)\right\}^{\frac{1}{2}}} dz,$$

where the k_l are the characteristic values of η, while the path of integration γ proceeds from r through the lower half of the complex plane to a point on the real axis preceding all k_l and from there returns to r through the upper half-plane.

In testing for the presence or absence of serial correlation our regression η is the ratio of products of successive observations, and $k_l = \sigma^2 \cos l\pi/(T + 1)$. Replacing this set of discrete values in the above integral by a continuous variable of similar distribution, the following approximation to the distribution of r is found:

$$h^*(r) = \frac{T - 2}{\pi} 2^{\frac{1}{2}T} \int_{\arcsin r}^{\frac{\pi}{2}} (\sin\phi - r)^{\frac{1}{2}T - 2} \sin\left(\frac{T}{4}\pi - \frac{T + 1}{2}\phi\right) \cos^{\frac{1}{2}}\phi \, d\phi$$

CONSTITUTION
OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

ARTICLE I

Name and Purpose

1. This organization shall be known as the Institute of Mathematical Statistics.
2. Its object shall be to promote the interests of mathematical statistics.

ARTICLE II

Membership

1. The membership of the Institute shall consist of Members, Fellows, Honorary Members, and Sustaining Members.
2. Voting members of the Institute shall be (a) the Fellows, and (b) all others who have been members for twenty-three months prior to the date of voting.

ARTICLE III

Officers, Board of Directors, Committee on Membership, and Committee on Publications

1. The Officers of the Institute shall be a President, two Vice-Presidents, and a Secretary-Treasurer, elected for a term of one year by a majority ballot at the annual meeting of the Institute. Voting may be in person or by mail.

(a) Exception. The first group of Officers shall be elected by a majority vote of the individuals present at the organization meeting, and shall serve until December 31, 1936.

2. The Board of Directors of the Institute shall consist of the Officers and the previous President.

3. The Institute shall have a Committee on Membership composed of three Fellows. At their first meeting subsequent to the adoption of this Constitution, the Board of Directors shall elect three members as Fellows to serve as the Committee on Membership, one member of the Committee for a term of one year, another for a term of two years, and another for a term of three years. Thereafter the Board of Directors shall elect from among the Fellows one member annually at their first meeting after their election for a term of three years. The president shall designate one of the Vice-Presidents as Chairman of this Committee.

4. The Institute shall have a Committee on Publications composed of three Members or Fellows elected by the Board of Directors. The President shall designate a Vice-President as Ex Officio Chairman of this Committee.

ARTICLE IV

Meetings

1. A meeting for the presentation and discussion of papers, for the election of Officers, and for the transaction of other business of the Institute shall be held annually at such time as the Board of Directors may designate. Additional meetings may be called from

THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG J. NEYMAN

WITH THE COÖPERATION OF

H. C. CARVER
H. CRAMÉR
W. E. DEMING
G. DARMOIS
R. A. FISHER
T. C. FRY
H. HOTELLING
R. VON MISES
E. S. PEARSON
H. L. RIETZ
W. A. SHEWHART

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I-IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.